Journal of Proteomics 128 (2015) 306–320

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects

Bjoern Titz *[,1], Alain Sewer [1], Thomas Schneider [1], Ashraf Elamin [1], Florian Martin [1], Sophie Dijon, Karsta Luettich, Emmanuel Guedj, Gregory Vuillaume, Nikolai V. Ivanov, Michael J. Peck, Nveed I. Chaudhary, Julia Hoeng, Manuel C. Peitsch

Philip Morris International R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland

ARTICLE INFO

Article history:
Received 22 May 2015
Accepted 15 August 2015
Available online 22 August 2015

Keywords:
COPD
Smoking
Sputum
Systems toxicology
Quantitative proteomics

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is one of the most prevalent lung diseases. Cigarette smoking is the main risk factor for COPD. In this parallel-group clinical study we investigated to what extent the transitions in a chronic-exposure-to-disease model are reflected in the proteome and cellular transcriptome of induced sputum samples. We selected 60 age- and gender-matched individuals for each of the four study groups: current asymptomatic smokers, smokers with early stage COPD, former smokers, and never smokers. The cell-free sputum supernatant was analyzed by quantitative proteomics and the cellular mRNA fraction by gene expression profiling. The sputum proteome of current smokers clearly reflected the common physiological responses to smoke exposure, including alterations in mucin/trefoil proteins and a prominent xenobiotic/oxidative stress response. The latter response also was observed in the transcriptome, which additionally demonstrated an immune-cell polarization change. The former smoker group showed nearly complete attenuation of these biological effects. Thirteen differentially abundant proteins between the COPD and the asymptomatic smoker group were identified including TIMP1, APOA1, C6orf58, and BPIFB1 (LPLUNC1). In summary, our study demonstrates that sputum profiling can capture the complex and reversible physiological response to cigarette smoke exposure, which appears to be only slightly modulated in early-stage COPD.

1. Introduction

The prevalence of chronic obstructive pulmonary disease (COPD), a common lung disease, is estimated to be between 8% and 10% of the adult population in developed countries [1], and the World Health Organization (WHO) identified COPD as the third most common cause of death in 2012 [2]. According to the chronic obstructive lung disease diagnosis and treatment guidelines (GOLD; http://www.goldcopd.com), COPD is characterized as a lung disease with persistent airflow limitation, which is usually progressive and incorporates both emphysema and chronic bronchitis [3]. However, because of its complexity and variability, it has been proposed recently that COPD is best regarded as a syndrome rather than a single disease [4]. In the developed world, cigarette smoking has been identified as the main risk factor for COPD development and its progression [5–8]. For example, in a 25-year follow-up study in Denmark it was found that the risk for continuous smokers to develop COPD was at least 25% [9]. In addition to smoking, other risk factors have been identified, including other sources of exposure (e.g., occupational exposure or air pollution) and genetic risk factors (e.g., α1-antitrypsin deficiency) [1,10,11].

Initiation of COPD is only incompletely understood, but is clearly facilitated by smoke exposure [12–14]. In asymptomatic smokers, the response of the lungs to smoke exposure includes an elevated oxidative stress response, inflammation, and an increased mucus production [15]. However, the lungs remain in homeostasis with, for example, sufficient redox capacity to deal with the oxidative stress and the ability to keep the inflammatory response in check [16]. During development of COPD, these normal, physiological, and inflammatory responses of the lungs appear to be amplified, go further out of balance, and consequently result in more permanent chronic inflammation and structural lung damage [3]. Consistent with this, it has been argued that the manifestation of COPD in long-term smokers represents more a quantitative rather than qualitative difference in the underlying biological effects [17]; i.e., at some point the physiological balancing mechanisms are at their limits and permanent structural degradation follows. The exact transition point/threshold will likely be difficult to define, but can be expected to depend on general genetic susceptibilities and on time-dependent processes that affect the coping potential of the tissue, such as development of cellular senescence [18,19]. These points can be summarized in

* Corresponding author.
E-mail address: bjorn.titz@pmi.com (B. Titz).
[1] Contributed equally.

http://dx.doi.org/10.1016/j.jprot.2015.08.009

a basic toxicant-exposure-to-disease-transition model for COPD (see Section 4).

Several biological matrices have been explored to investigate effects of respiratory toxicants and diseases on the human lung. The sampling sources ranged from induced sputum to bronchoalveolar lavage fluid to lung biopsies [20,21] and, of these, induced sputum has the advantage of being the least invasive and best tolerated technique, allowing for its safe application in large study cohorts. Sputum is induced by inhalation of hypertonic saline, and subsequently coughed up and collected for analysis. Previous respiratory toxicology and lung disease studies investigated alterations in the sputum proteome. The first, published by Nicholas et al. [22], reported the protein components of a single smoker's sputum using two-dimensional (2D) gel and LC–MS/MS analyses. In a subsequent study, Nicholas et al. [23] analyzed a larger cohort of COPD subjects (GOLD stages 1–3) and asymptomatic smokers using a 2D gel approach. Gray et al. [24] employed SELDI-TOF for the identification of potential biomarkers for asthma, cystic fibrosis, and COPD. Ohlmeier et al. [25] analyzed seven non-smokers, asymptomatic smokers, and COPD subjects with 2D difference gel electrophoresis and identified elevated polymeric immunoglobulin receptor (PIGR) levels in COPD subjects compared with asymptomatic smokers. In contrast, the number of reported sputum transcriptomics studies is limited. For example, Singh et al. [26] analyzed the sputum transcriptome profiles of former smokers with COPD (GOLD stages 2–4) from the ECLIPSE study – a non-interventional, observational, multicenter, three-year study in subjects with COPD [27] – and identified significant changes in genes associated with the severity of airflow limitation and emphysema.

Here, we present results from a clinical case–control study with 60 age- and gender-matched individuals for each of the four study groups: current asymptomatic smokers, smokers with early stage COPD, former smokers, and never smokers (Fig. 1A). Specifically, in this manuscript we focus on the question how cigarette smoke exposure, smoking cessation, and the presence of early-stage COPD are reflected in the proteome and cellular transcriptome of induced sputum samples.

2. Materials and methods

2.1. Subjects

In this study, we used a parallel-group case-controlled study design to determine changes in the sputum proteome and transcriptome of smokers with COPD, asymptomatic (no COPD) smokers (CS), asymptomatic (no COPD) former smokers (FS), and asymptomatic never smokers (NS). The study was conducted between July 2011 and December 2012 at a single clinical site in London, UK, after approval from the National Health Service (NHS) Black County Ethics Committee and in strict compliance with the International Conference on Harmonisation—Good Clinical Practice (ICH-GCP) guidelines. The study has been registered on ClinicalTrials.gov with identifier NCT01780298.

Male and female subjects aged between 41 and 70 years were enrolled with a completed total subject number of 60 per group. If a subject discontinued participation (for medical or personal reasons), he/she was replaced. During the course of the study, sputum, blood, and nasal samples were collected and a number of physiological and clinical measurements were recorded from a total of 240 subjects. Here, we report on the findings for the sputum proteome and transcriptome. Subjects in each of the three control groups (CS, FS, and NS) were matched to subjects in the COPD group by age (±5 years), ethnicity, and gender, and all smoking subjects had a smoking history of at least 10 pack-years. For this purpose, a match ID was defined for each paired group of four subjects.

2.2. Sputum collection

Sputum induction was performed at visit 1 (screening), visit 1a (at the discretion of the investigator), and again at visits 2 and 4. At visit 1 (and visit 1a), subjects underwent the sputum induction procedure to confirm that they were able to produce a sputum sample weighing at least 0.1 g; this was an eligibility criterion for the study. Subjects unable to produce an adequate induced sputum sample at visit 2 repeated the procedure at visit 3.

All subjects with a forced expiratory volume in one second (FEV_1) of ≥80% predicted and with no respiratory disease performed the sputum induction procedure directly. Subjects with COPD or an FEV_1 of ≤80% predicted were administered 200 μg of salbutamol by metered dose inhaler. The FEV_1 was then assessed 15 min after salbutamol administration and subjects with a highest FEV_1 value of ≥1.0 L or ≥50% predicted could proceed to sputum induction.

Prior to sputum induction, the subject was asked to blow his/her nose. Nasal passages were closed with a soft nose-clip and the subject instructed to inhale 3% hypertonic sodium chloride solution over 5 min. After inhalation or if able to expectorate before 5 min had passed, the subject was asked to blow his/her nose, gargle, and rinse their mouth using room temperature water. The subject was then instructed to cough up sputum, which was collected in a pre-labeled cup. Where possible the subject was to complete three cycles of sputum expectoration using 3%, 4%, and 5% hypertonic sodium chloride. Sputum from all three cycles was collected in the same cup. If a subject's FEV_1 fell ≥20% of the best post-bronchodilator baseline value or if significant symptoms were seen, sputum induction was stopped and the subject given bronchodilator therapy as needed. Sputum samples were placed on ice and processed within 2 h. Sputum plugs were selected and solubilized in DTT, and the cell phase was collected cytometry and transcriptomics evaluation by centrifugation while the supernatant was collected in separate cryovials and stored at −80 °C until subsequent proteomic analysis.

2.3. Proteomics analysis

Sputum samples from all 240 study subjects were processed in random order. A reference sample was included by mixing equal protein amounts of all 240 samples. Proteins were extracted by acetone precipitation and desalted, followed by trypsin digestion. The tryptic peptides were labeled using Tandem Mass Tag™ 6-plex (TMTsixplex™) reagents (Thermo Scientific, Waltham, MA, USA), and one labeled sample each representing the four study groups plus the reference sample were mixed in equal protein amounts. The paired subjects in each TMT set were matched to the COPD sample by age, ethnicity, and gender (and a match ID was assigned to each set). These labeled mixes were then purified to remove unincorporated TMT™ reagents and subsequently analyzed by liquid chromatography tandem mass-spectrometry (LC–MS/MS) using an EASY-nanoLC 1000 instrument connected online to a Q Exactive™ mass-analyzer (Thermo Scientific). Peptides were fractionated on a 50 cm C18RP RSLC EASY-spray™ column (2 μm particle size; Thermo Scientific) at a flow rate of 200 nL/min with a 200 min gradient from nanoLC buffer A (5% acetonitrile, 0.2% formic acid) to 40% acetonitrile, 0.2% formic acid. Each sample set was analyzed in duplicate in fast and sensitive analysis mode as described previously [28]. LC–MS/MS data from both injections were merged and compared against the human reference proteome in the UniProt database (http://www.uniprot.org/, version January 2014). Proteome Discoverer™ 1.4 (Thermo Scientific) was used for the database searches with the SEQUEST™ HT and Mascot® 2.4 search algorithms. The Percolator node of Proteome Discoverer™ was used to estimate peptide-level adjusted p-values (q-values), and the peptides were filtered for a q-value <0.05 (i.e., the false discovery rate (FDR) was controlled at the 5% level). The quantification of TMT™ reporter ions and the peptide-to-protein (group) assignments was performed with Proteome Discoverer™. TMT™ peptide-level data were exported and further processed in the R statistical environment [29]. Quantitative data were filtered for "unique" quantitative results; e.g., by removing redundant results from multiple search engines. To improve data quality, multiplexed

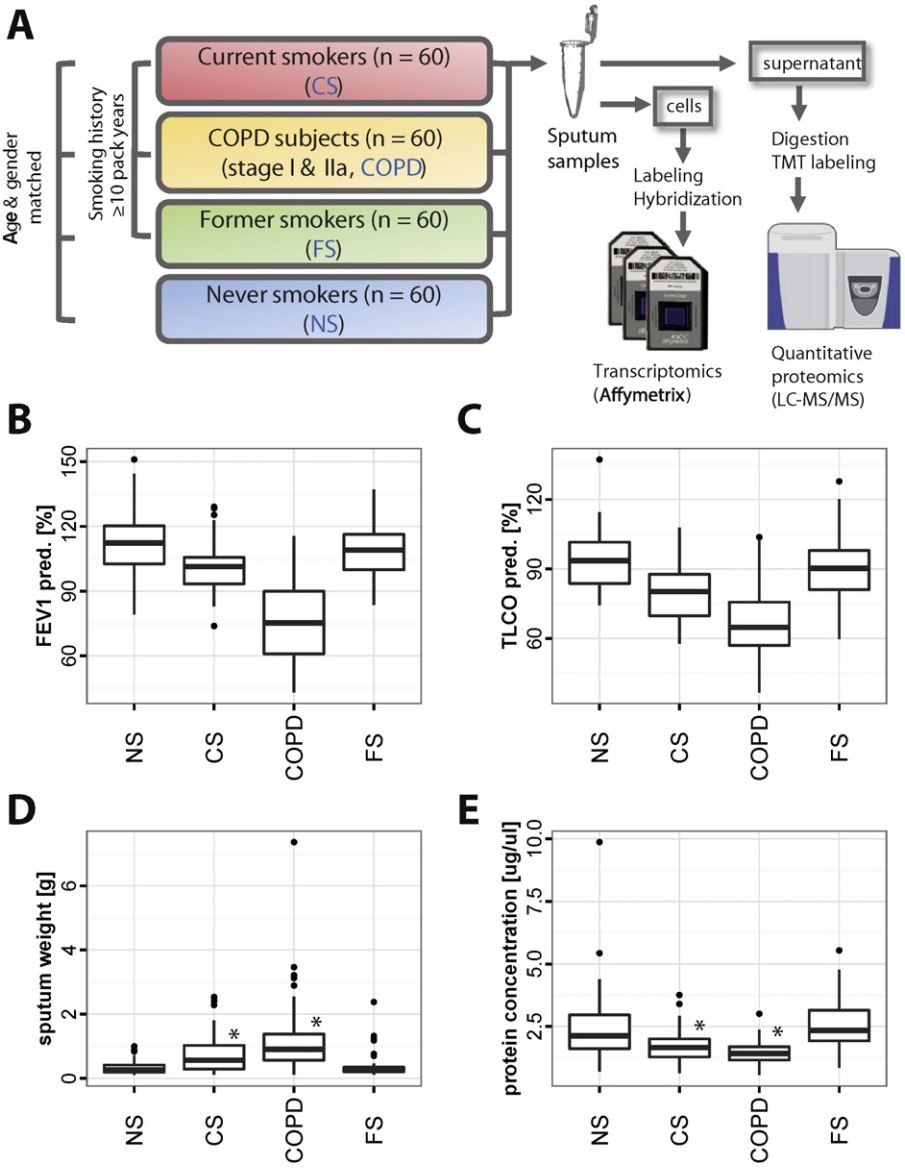


Fig. 1. Parallel-group case-controlled study design to determine proteome and transcriptome differences in COPD subjects, asymptomatic smokers, asymptomatic former smokers, and asymptomatic never-smokers. (A) Schematic representation of the study design. Induced sputum samples were collected from 60 individuals in each study group (current smokers [CS], stages I and IIa COPD subjects, former smokers [FS], and never smokers [NS]). COPD subjects were current smokers. Subjects in the COPD, CS, and FS groups had a smoking history of ≥10 pack years. All groups were age and gender matched. Sputum samples were divided into the cellular fraction for transcriptomic analysis and the cell-free supernatant for quantitative proteomic analysis. (B) Boxplot showing the median and quartile ranges of the lung function parameter FEV_1 predicted for all study groups. (C) As in B, for the carbon monoxide transfer factor (TLCO predicted). (D) Boxplot for the measured weights of the induced sputum samples. * indicates the p-values (Welch t-test) compared with NS were <0.05. (E) Boxplot for the measured protein concentrations for the sputum supernatants (measured after protein precipitation). * indicates the p-values (Welch t-test) compared with NS were <0.05.

sample runs with median log parent intensity below the 25%ile–1.5 IQR (interquartile range) or above the 75%ile + 1.5 IQR were classified as outliers. With this, 216 out of 240 samples were retained for the subsequent analysis (54 of each group). Global variance stabilizing normalization was performed [30,31]. Each TMT™ reporter ion set was normalized to its median, and protein expression values were calculated as the median of these normalized peptide-level quantitation values [32]. Only proteins quantified for at least 2/3 of the samples of each study group were considered for differential abundance analysis. The four subject groups (NS, CS, FS, and COPD) enabled six pairwise comparisons, for which the statistics of differential protein expressions were calculated using the *limma* package [33]. For each comparison, a linear model was fitted to the expression data of the two considered subject groups only, which further included a covariate variable (match ID) taking into account the 60 subject strata defined in the study design. The Benjamini–Hochberg (BH) FDR method was used to correct for multiple testing effects [33]. Proteins with BH-adjusted p-values <0.05 were considered as significantly differentially abundant. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [34] via the PRIDE partner repository with the dataset identifier PXD001977. Note that the data for the full set of 240 samples is provided and the 216 samples considered for the analysis presented here are indicated in the "Comments [IncludedForAnalysis]" column in the sample annotation file.

2.4. Generation of transcriptomics data

RNA was extracted and processed by AROS Applied Biotechnology A/S (Aarhus, Denmark) and hybridized to Affymetrix Human Genome U133 Plus 2.0 Arrays according to standard operating procedures. 192

out of 240 microarrays (CEL files) passed initial QC criteria and were subjected to further data analysis in the R statistical environment [29]. Raw RNA expression data were first preprocessed using the *affy* and *gcrma* packages of the Bioconductor suite of microarray analysis tools [33,35,36]. GeneChip Robust Multiarray Average (GCRMA) background correction and quantile normalization were used to generate probe set-level expression values [37]. Quality controls were then iteratively performed using the *arrayQualityMetrics* package [38] to retain 179 valid arrays. As for the proteomics analysis, the four subject groups (NS, CS, FS, and COPD) enabled six pairwise comparisons, for which the statistics of probe set-level differential expressions were calculated using the *limma* package [33]. For each comparison, a linear model was fitted to the expression data of the two considered subject groups only, which further included a covariate variable (match ID) taking into account the 60 subject strata defined in the study design. Probe set-level results were attributed to the 23.414 HGNC gene symbols underlying our network collection (see below) using the best probe set-gene symbol annotation determined by the overall lowest p-values in the six pairwise comparisons. Raw p-values were then adjusted for multiple testing using the Benjamini–Hochberg procedure [33]. Microarray data were submitted to ArrayExpress with the accession number E-MTAB-3604.

2.5. Identification of functional clusters

Functional association networks were derived for the differentially abundant proteins or genes and functional association clusters were identified to guide functional interpretation. The functional associations between genes or proteins were obtained from the STRING database (version 9.1) [39]. To identify functional protein clusters associated with the current smoking status, functional associations with at least medium confidence (score >0.4) between the differentially up- (or down-) regulated proteins in all smoker vs. non-smoker comparisons (CS vs. NS, COPD vs. NS, CS vs. FS, COPD vs. FS) were extracted and clusters were identified using an information theoretic approach for community identification [40]. Functional annotation of the clusters was guided by enrichment analysis with the TOPPGene tool [41]. To allow for summarization of the transcriptomics data with a larger number of differentially expressed mRNAs as a concise and interpretable association network, we used the dnet approach [42]. With this, a concise gene association network of about 100 nodes with maximized overall differential expression for CS vs. NS was identified. Cluster identification and annotation was performed as for the proteomics data. Network analysis and visualization was supported by the igraph package in R [43,44].

2.6. Network analysis for transcriptomics data

In order to interpret the transcriptomics results, we also applied a systems biology approach called “Network Perturbation Amplitude” (NPA), which uses the differential expression of predefined sets of genes (transcriptional layer) to deduce the changes in the activities of the upstream biological mechanisms that collectively affect these genes (functional layer) [45,46]. These mechanisms have been assembled as so called backbone nodes into causal network models describing specific processes such as cellular stress and inflammatory responses [47,48]. One advantage of the NPA approach is that the calculated perturbation amplitudes can be quantitatively contrasted across multiple pairwise comparisons, very much like gene differential expression, but at a higher level of the systems organization (backbone node or even full network instead of single genes).

Because sputum samples contain multiple cell types (mostly macrophages and neutrophils) and because the relative proportions of these multiple cell types can vary between the four subject groups (NS, CS, FS, and COPD), care is needed when interpreting the changes in the sputum cellular transcriptomes. In particular, this is of concern when applying the cell type-specific inflammatory network models [47], which implicitly assume that the changes in the cellular mRNA can be attributed to the differential regulation within one cell type. Thus, we calculated the NPA values at the level of the individual backbone nodes in a network model-independent manner by applying the so called *strength* metric. Comparing these NPA results and gene expression changes for our study with published datasets on the gene expression response of macrophages (see below) enabled us to determine which cell type-specific inflammatory network models were applicable for a given pairwise comparison. The appropriate models were then used to identify biological processes that likely caused the changes observed in the sputum transcriptomes.

The public gene expression datasets (Supplementary Table S3) were downloaded from the Gene Omnibus (GEO) database and preprocessed as described above for our data. The downstream genes (i.e., the transcriptional layer) that are connected to the backbone nodes of the cellular stress and inflammation network models used in the NPA calculations were extracted from the Selventa Knowledgebase, which is a comprehensive repository containing over 1.5 million nodes and over 7.5 million edges [45]. The Selventa Knowledgebase is derived from peer-reviewed scientific literature as well as other public and proprietary databases — a part of which is publicly available through the openBEL portal [www.openbel.org]. The GSEA calculations were run on a local installation of the analysis software [49].

2.7. Identification of discriminant predictive proteins

A discriminant set of proteins for a given comparison (e.g., CS vs. NS) were derived either by logistic lasso regression [50] or by linear discriminant analysis (LDA) with a forward selection procedure [51]. The forward selection is aiming at first finding the most discriminant protein (by leave-one-out cross-validation) by testing all the available ones using LDA, then subsequently test for the second best (together with the first selected one) and so on, until having included N proteins, where N is specified by the user based on the expected complexity of the model. The regularization parameter for the logistic lasso regression was selected by 10-fold cross-validation. The performance of each procedure was evaluated by k-fold cross validation, repeated L times where $k = 10$ and $L = 5$.

3. Results

3.1. Study population and sputum sampling

To assess how smoking and mild COPD disease are reflected in the sputum proteome and transcriptome, we collected induced sputum from four study groups: early-stage COPD subjects (GOLD stage 1–2 and current smokers) (COPD), current smokers without COPD disease (CS), never smokers (NS), and former smokers (FS). The rationale was to cover all relevant transitions during the early course of the disease. Each cohort ($n = 60$) was matched by age (within 5 years) and gender, and individuals were required to have a smoking history greater than 10-pack years (except for never smokers). Former smokers had quit for at least 1 year prior to the start of the study (Fig. 1A, Table 1). Importantly, our selection criteria were different from other related studies such as the ECLIPSE study [27,52]: we selected only subjects with mild and moderate COPD (29 GOLD stage 1, 31 GOLD stage 2) and subjects with recent infections or a history of exacerbations of COPD were excluded. FEV_1% predicted and the transfer factor for carbon monoxide (TLCO predicted) exemplify the decline in lung function in the COPD cohort (Fig. 1B/C). The exclusion of subjects with a recent history of infections or exacerbations of COPD likely explains why we did not observe the previously reported increase in neutrophil percentages for the COPD and CS groups in our study population (Supplementary Figure S1) [53,54]. However, recent reports have indicated that this association may be less consistent and possibly more tightly linked to a current active inflammatory process rather than COPD as a whole [55].

The sputum sampling revealed basic differences between the study groups. As expected, more sputum was obtained from current smokers (CS and COPD), especially COPD subjects, than from NS and FS (Fig. 1D). Conversely, sputum proteins were more diluted, when sampled from CS and COPD than from NS and FS groups (Fig. 1E).

3.2. Sputum proteome and transcriptome reflect smoking and COPD status

Induced sputum samples were collected from all 240 study participants, the cell-free supernatant was subjected to TMT™-based quantitative proteomic analysis, and the cellular sputum fraction was subjected to microarray-based transcriptomic analysis. For proteomic analysis, 54 subjects in each group remained after quality filtering (i.e., 216 in total) (see Methods). We used a normalization procedure that corrects for the observed intensity difference between the sample groups to focus on the protein abundance differences within the sputum proteome fraction accessible to mass-spectrometry analysis. For the transcriptomic analysis, 179 samples remained after quality filtering.

Differentially abundant sputum proteins and mRNAs between the study groups were identified (FDR-adjusted p-value <0.05). Volcano plots of effect size and significance showed clear differences between the smoker (CS and COPD) and non-smoker (FS and NS) groups for both the proteomics and the transcriptomics data (Fig. 2A/B, Supplementary Table S1). Among the affected proteins, aldehyde dehydrogenase 3A1 (ALDH3A1) showed the strongest and most significant increase in the smoker groups (CS and COPD) and, strikingly, was by itself sufficient to accurately predict the current smoking status of the study subjects (Supplementary Fig. 2). In contrast, the FS and NS groups demonstrated limited statistically significant differences in protein expression with only the S100 calcium binding protein A6 (S100A6) showing significantly lower abundance in FS than in NS. No differentially expressed mRNAs between the early-stage COPD and the CS groups were identified. However, 13 proteins were differentially abundant between these two groups, which provided the first evidence for COPD-specific effects in the induced sputum samples (discussed below).

Importantly, for both the proteomics and transcriptomics data, pairwise comparisons between the groups showed a high correspondence between all four current smoker vs. non-smoker group comparisons (CS vs. NS, CS vs. FS, COPD vs. NS, and COPD vs. FS) (Fig. 2C/D). With this, the current smoking status of the study subjects clearly dominated the observable effects in the induced sputum proteome and transcriptome.

3.3. Smoking is reflected by the activation of several compensatory mechanisms and a change in immune cell polarization

Several compensating mechanisms are activated in asymptomatic current smokers that allow maintenance of homeostasis and prevent manifestation of disease. To assess which of these mechanisms might be detectable in our induced sputum data, we identified functional clusters enriched for differentially abundant proteins and differentially expressed mRNAs (Fig. 3A and B, Supplementary Figure S3). To compensate for the lower number of proteins quantified by proteomics and to maximize sensitivity, we considered all differentially abundant proteins between the current smoker (CS/COPD) and current non-smoker (NS/FS) groups.

Several functional protein clusters affected by cigarette smoke exposure were identified in the sputum proteome (Fig. 3A). These included mucin/trefoil proteins (e.g., MUC5AC and TFF1/3), xenobiotic metabolism enzymes (e.g., ALDH3A1, NQO1, and GSTA1), peptidase regulators (e.g., TIMP1 and SERPINB1), and proteins involved in redox processes (e.g., TXN, PRDX1, and SOD1). Among the down-regulated clusters were likely blood plasma-derived proteins (e.g., ALB, APOA1, and TF) and immunoglobulins (e.g., IGHG1-4 and IGKC). With this, the observed effects in the sputum proteome reflect several of the main known effects of cigarette smoke exposure including the xenobiotic and oxidative stress response, changes in mucin production, and alterations in the protease balance [56–58].

The functional clusters identified for the differentially expressed mRNAs in sputum cells (CS vs. NS) included xenobiotic and oxidative stress (e.g., CYP1B1, NQO1, and GSR) and immune-related (e.g., CXCL10, CXCL11, and GBP4) gene clusters (Fig. 3B). Whereas the xenobiotic and oxidative stress cluster contained mainly up-regulated mRNAs and thus reflected a common exposure response, the mostly down-regulated immune-related cluster required further investigation. We noted that the most strongly enriched pathway for this down-regulated cluster was the interferon signaling pathway (Reactome database (http://www.reactome.org/), q-value (BH) = 5.8E−10, ToppGene tool [41]). In further support of the down-regulation of interferon signaling, several binding sites for interferon-regulatory factors and an interferon gamma gene set were significantly negatively enriched in the CS vs. NS comparison in a gene set enrichment analysis (Supplementary Table S2).

To better understand the above described exposure response of the sputum cell population, we compared the sputum transcriptomics data with five public data sets (Fig. 3C, Supplementary Table S3). These studies analyzed the smoke exposure response of alveolar macrophages and, interestingly, they all showed a positive correlation and a significant overlap of the differentially expressed genes with our study (Fisher's test p-value <0.05). This similarity suggested that macrophages were the main drivers of the changes observed in our sputum mRNA data and enabled us to employ a previously published causal macrophage activation network to better understand the response of these cell populations [47]. Except for one data set (GEO: GSE27002) all others demonstrated significant perturbation of the macrophage activation network as a whole (Supplementary Fig. 4 and 5) [45]. In addition, and further supporting our observation of altered IFN signaling in smokers compared to never-smokers, a major shared component of this network response was the down-regulation of the interferon

Table 1
Characteristics of study subjects.

	Early-stage COPD subjects	Asymptomatic smokers	Asymptomatic never smokers	Asymptomatic former smokers
Group label	COPD	CS	NS	FS
N (before/after filter)	60	60	60	60
Age (years)	57.17 (±7.16)	55.31 (±6.91)	55.46 (±7.45)	56.34 (±7.39)
Male:female ratio %	60:40	53:47	57:43	58:42
Pack years	44.0 (10.0–117.5)	25.5 (10.0–70.0)	0	31.0 (10.3–75.0)
Years in cessation	0	0	0	8.85 (1.06–44.3)
BMI kg/m^3	26.46 (±3.7)	27.33 (±3.56)	26.48 (±3.72)	27.33 (3 ± 3.56)
Alcohol units/week	6.69 (±7.66)	5.92 (±6.31)	5.78 (±6.71)	8.45 (±7.10)
FEV_1% predicted	75.28 (43.05–115.50)	101.35 (73.85–129.10)	112.35 (79.20–151.05)	109.05 (73.85–129.10)
Diastolic blood pressure (mm Hg)	73.5 (57–92)	71.0 (61–92)	72.0 (56–88)	72.0 (56–92)
Systolic blood pressure (mm Hg)	121.0 (90–148)	121.0 (93–147)	122.0 (94–179)	123.5 (101–151)
Heart rate (bpm)	70.0 (44–101)	68.0 (50–95)	62.0 (43–91)	66.0 (43–101)
Respiration rate (breaths per min)	15.0 (11–19)	15.0 (11–20)	16.0 (11–20)	15.0 (11–20)

FVC, forced vital capacity; FEV_1, forced expiratory volume in the first second; BMI, Body Mass Index; bpm, beats per minute. Data are presented as mean (±SEM) or as median (range).

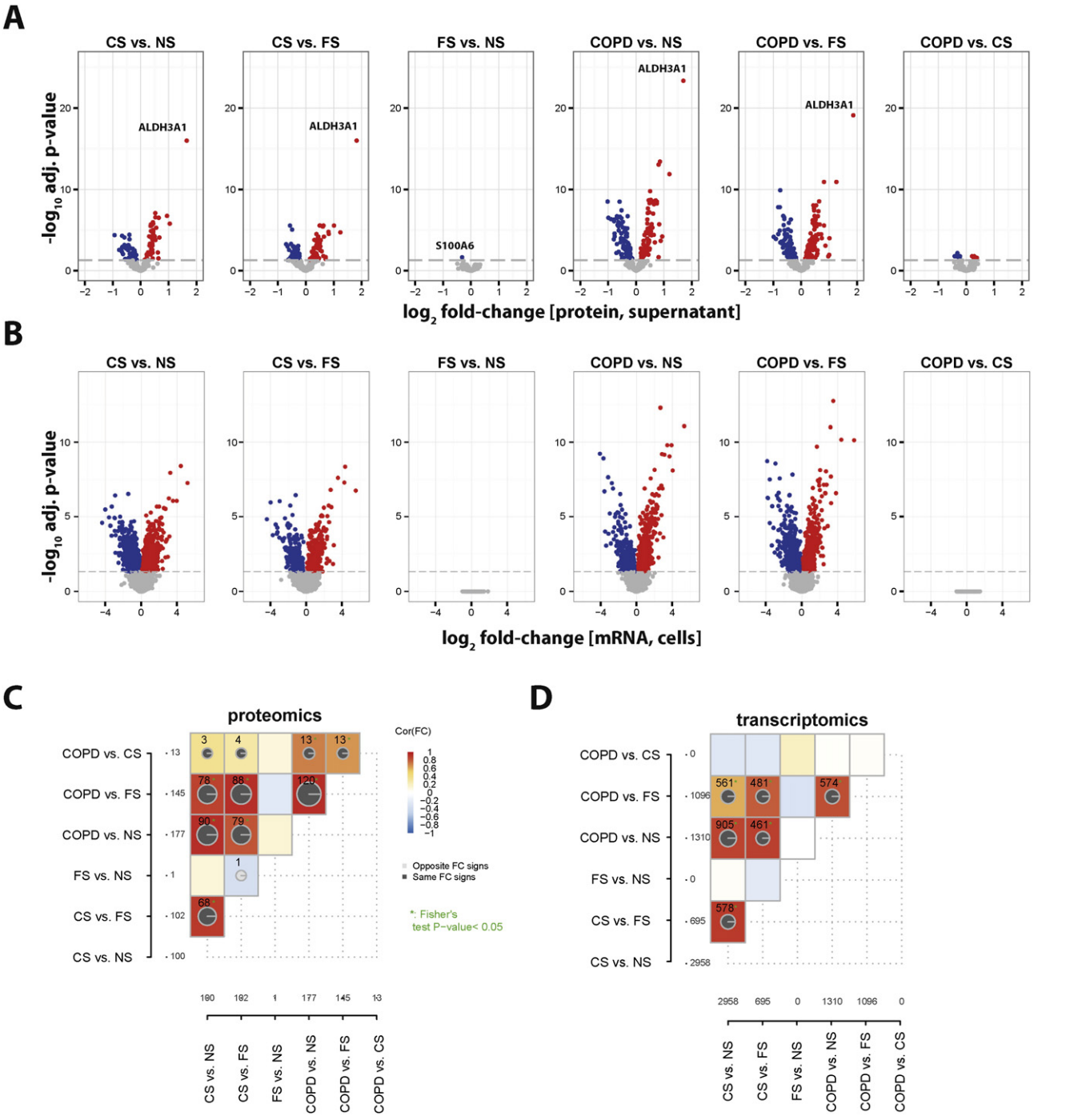


Fig. 2. Sputum proteome and transcriptome reflect the smoking status. (A) Volcano plots showing the magnitude (log2 fold-change) and significance (−log10 Benjamini–Hochberg adjusted p-value) of differential sputum protein expression for the six indicated group comparisons. Significantly up-regulated proteins are marked in red, significantly down-regulated proteins are marked in blue (adjusted p-value <0.05). Note that 13 proteins were differentially abundant between the COPD and CS groups. (B) Volcano plots showing the differential mRNA expression in sputum cells. Significantly up-regulated mRNAs are marked in red, significantly down-regulated mRNAs are marked in blue (adjusted p-value <0.05). (C) Comparison chart for the differentially abundant sputum proteins. The correlation coefficient for the comparisons is color-coded, and the number of shared differentially expressed proteins is indicated (total numbers in the margins). The pie charts show the percentage of shared differentially abundant proteins with the same direction of fold change (FC sign). The green * indicates the observed overlap of the differentially abundant proteins was significant. (D) Comparison chart for the differentially expressed mRNAs. Other details are the same as those in (C).

gamma node (IFNG) (Fig. 3D, Supplementary Fig. 6). In contrast, network analysis supported up-regulation of STAT3 activity upon smoke exposure.

Strikingly, IFNG and STAT3 have been associated with different macrophage polarization states: IFNG with M1 and STAT3 with M2 polarization [59] and macrophage polarization changes upon smoke exposure have already been reported for the compared studies [60,61]. Specifically, Shaykhiev et al. (GEO: GSE13896) found that alveolar macrophages of smokers "*exhibited a unique polarization pattern characterized by substantial suppression of M1-related inflammatory/immune genes and induction of genes associated with various M2-polarization programs*" [60]. When we directly assessed the behavior of the M1- and M2-phenotype-related genes defined by Shaykhiev et al. [60], we found a surprisingly high correspondence between the response of

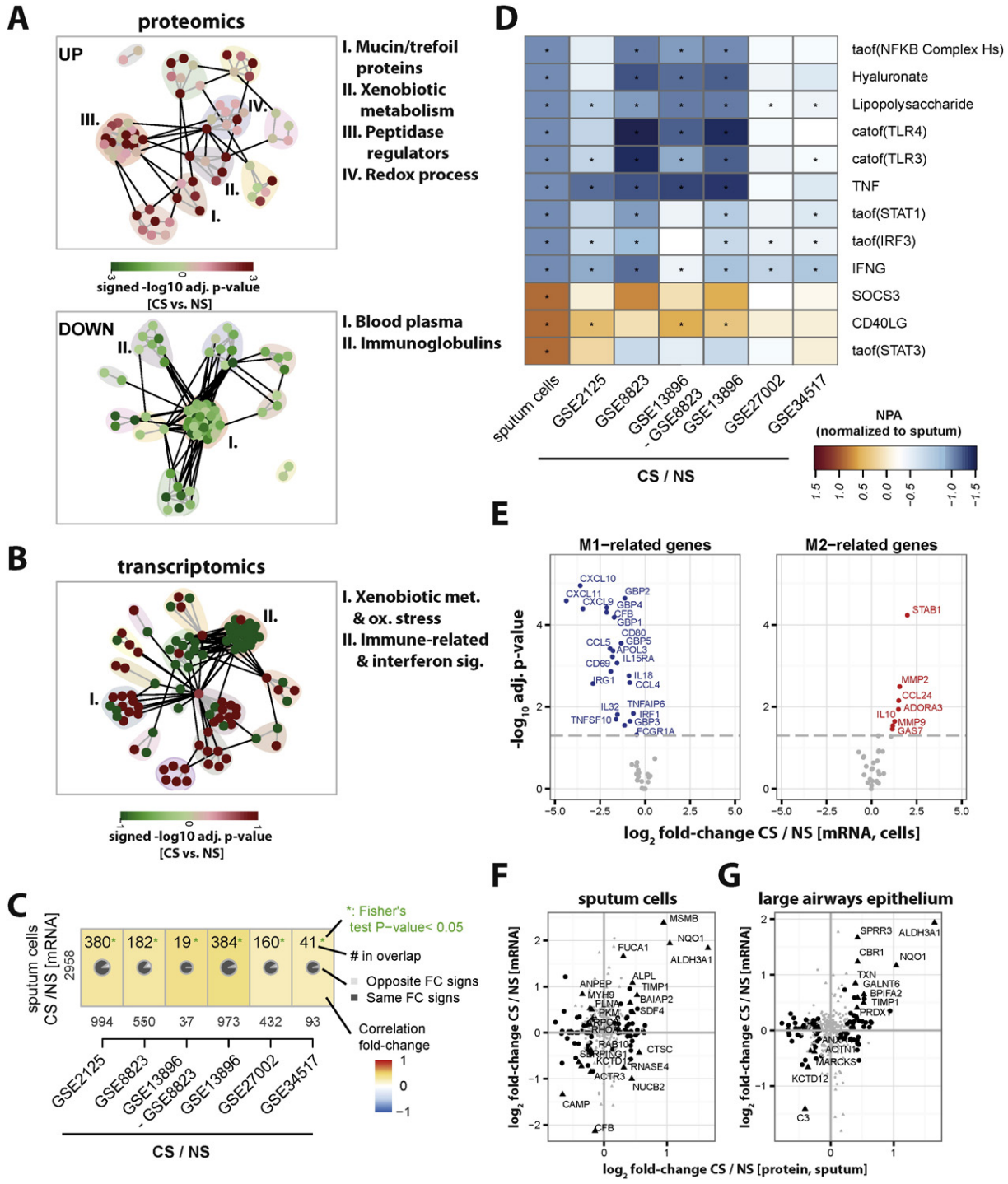


Fig. 3. Sputum proteome and transcriptome reveal biological effects of cigarette smoke exposure. (A) Functional protein clusters identified for the main smoke exposure effect by sputum proteomics. A functional association database (STRING database [39]) was queried with all significantly up- (top panel) or down- (bottom panel) regulated proteins in the smoker (COPD or CS) vs. non-smoker (NS or FS) groups. The identified association networks were clustered to guide functional interpretation. The identified clusters are marked and those with a clear biological function are annotated. The color of the nodes (proteins) represents the signed $-\log_{10}$ adjusted p-value of the CS vs. NS group comparison. (B) Functional gene clusters identified for the main smoke exposure effect by sputum transcriptomics. The dnet algorithm was used to identify an "activated" functional association subnetwork for the CS vs. NS comparison [42]. Other details are similar to (A). (C) Comparison of mRNA response in sputum samples (CS vs. NS) and five alveolar macrophage exposure studies (Supplementary Table S3). The CS vs. NS effect on the transcriptome is compared. The Pearson correlation coefficient for the observed fold-changes is color-coded; the number of detected and shared differentially expressed genes is indicated together with the fraction of fold-changes that show the same direction. (D) Network perturbation amplitude (NPA) values for the backbone nodes of the macrophage activation network. Nodes with a significant NPA value in the sputum cell CS vs. NS comparison are shown (rows). The NPA for these nodes is compared across studies (columns, see color key) and rescaled by the value for sputum cells. "*" denotes significance; p-value <0.05 for both *uncertainty* and *specificity* statistics [45]. (E) Volcano plots showing the cellular mRNA expression changes in the CS vs. NS groups for the M1- and M2-related genes as defined by Shaykhiev et al. [60]. Significantly up-regulated genes are marked in red, significantly down-regulated genes are marked in blue (adjusted p-value <0.05). (F) Fold-change comparison between the quantified sputum proteins and cellular sputum mRNAs (CS vs. NS). Significant differential protein abundance for CS vs. NS is indicated by large black dots (otherwise small gray dots). Significant differential mRNA expression for CS vs. NS is indicated by black triangles. Proteins/mRNAs with significant differential changes in both data sets are labeled. (G) Fold-change comparison between differentially abundant proteins in sputum and differentially expressed mRNAs in the expression data set for the large airways epithelium (both current smokers vs. non-smokers, GEO: GSE10135[62]). Other details are similar to those in (E).

alveolar macrophages and the arguably different cell population sampled by sputum induction (Fig. 3D, Supplementary Fig. 7); i.e., based on the M1/M2-signature genes the sputum cell population in current vs. never smokers clearly shifted from an M1 to an M2 phenotype. Importantly, these results also confirmed the relevance of macrophage-based mechanisms for the interpretation of the cigarette smoke exposure effect and the relevance of the "macrophage activation" model in our network analysis [47].

Finally, we asked how the observed differences in protein abundance relate to changes in mRNA expression. Because we sampled different sputum fractions for proteomics (supernatant) and transcriptomics (cellular fraction), we expected a limited overlap (Fig. 3F). However, in line with the functional cluster analysis, the genes encoding the main up-regulated oxidative/xenobiotic stress response proteins that were detected in the cellular fraction (ALDH3A1 and NQO1) were also found in sputum supernatant. Sputum proteins are derived from different sources including from secretory cells in the epithelium. Thus, we compared the significantly differentially abundant sputum proteins with differentially expressed genes in the large airway epithelium of smokers (GEO: GSE10135[62]) and found that five of the nine shared up-regulated proteins were oxidative/xenobiotic stress response proteins (e.g., ALDH3A1, NQO1, and TXN) (Fig. 3G). Another example of a shared up-regulated sputum protein was SPRR3, which is a component of the cornified cell envelope of stratified squamous epithelia [63]. With this, increased abundance of SPRR3 likely reflects the development of squamous metaplasia in smokers [64,65].

In summary, the sputum proteome and transcriptome reflected several of the biological effects of cigarette smoke exposure. Induction of the xenobiotic/oxidative stress response was shared between all compartments. In addition, the sputum proteome reflected alterations in mucus production and protease regulation and the transcriptome indicated a polarization of macrophages from the M1 toward the M2 state in the sputum cell population.

3.4. Long-term smoking cessation results in attenuation of exposure effects in sputum

Smoking cessation is the most effective measure to prevent COPD and to slow its progression [8,13]. It is known that upon smoking cessation the majority of observable exposure effects return to baseline levels: The oxidative stress response reverts within a year [66], modulation of the inflammatory state in the lung reverts on a similar time-scale [67], but an increased lung cancer risk is detectable for decades after smoking cessation [68]. In this study, subjects in the FS group had quit smoking for at least 12 months prior to the start of the study, and the majority (approximately 78%) had quit for more than five years. Thus, we asked to what extent the observed cigarette smoke exposure-related changes were still detectable in this long-term cessation group.

We detected only one differentially abundant protein (S100A6) and no differentially expressed transcripts between the FS and NS groups (Fig. 2A and B). Since this assessment depended on the chosen p-value threshold, we complemented it with a direct comparison of the observed fold-changes for the CS vs. NS and FS vs. NS comparisons (Fig. 4A and B). We expected a slope close to zero if the exposure effects in former smokers largely approached non-smoker levels, and this was indeed the case for both the proteomics and transcriptomics data (although a slightly increased slope was still observed for the proteomics data set). For example, the two highest up-regulated (xenobiotic/oxidative stress) proteins in CS vs. NS, ALDH3A1 and NQO1, exhibited lower levels in the FS than in the NS group.

When the identified biological effects in FS relative to NS were compared with those in CS relative to NS, a similar picture for both data types was obtained (Fig. 4C). The observed changes in the current smoker proteome including the mucin/trefoil proteins, the xenobiotic/oxidative stress response proteins, and the peptidase regulator cluster largely approached NS levels in the FS group. Interestingly, in the transcriptomics analysis the (interferon-related) immune-response cluster even demonstrated a slight up-regulation in the FS vs. NS comparison. This observation was further corroborated by the comparison of network perturbation amplitudes, which demonstrated an overall deactivation of the xenobiotic metabolism response, but a partial inversion of macrophage activation for FS vs. NS compared to CS vs. NS (Fig. 4D). For example, FS vs. NS showed an increase rather than a decrease in the NPA of the IFNG and the NFkB complex nodes (Fig. 4E), which for IFNG signaling was further supported by gene-set enrichment analysis (Supplementary Fig. 6). IFNG and NFkB signaling have been associated with M1 polarization [59] and – together with the gene set enrichment results for the M1-associated gene set (Supplementary Fig. 7) – this supports a persistent (but inverted) effect on the polarization of macrophages toward a M1 phenotype in the FS sputum.

In conclusion, the characteristics of both the sputum proteome and transcriptome of former smokers largely approached those seen in never-smokers. Nevertheless, some long-term effects of cigarette smoke exposure remained noticeable in former smokers' sputum as indicated by the increase in IFNG and NFkB signaling, which are both associated with a M1 polarization of the FS sputum cell population.

3.5. Early-stage COPD subjects demonstrate quantitative differences in sputum

A main question of this study was whether sputum samples reflect airway changes associated with early stages of COPD. A total of 13 sputum proteins demonstrated differential abundance between subjects with early-stage COPD and those that were asymptomatic (FDR-adjusted p-value <0.05, Fig. 5A). These 13 proteins showed a reasonable correlation with the FEV_1% predicted, which is in line with previous studies [23], and the carbon monoxide transfer factor (TLCO predicted) (Fig. 5B). The differential abundance of several of these 13 proteins has been linked previously to COPD or other lung diseases (Table 2); for example, increased abundance of TIMP1 (metalloproteinase inhibitor 1) in sputum [69,70] and decreased abundance of AHSG (alpha-2-HS-glycoprotein) [23] and ALB (serum albumin) [23,71] have been reported.

To expand the data comparison, we also evaluated the observed trends for the COPD-associated sputum proteins reported by Nicholas et al. [23] (Supplementary Fig. 8). Although they did not reach statistical significance in our study, all the proteins showed a consistent trend for the COPD subjects. For example, lipocalin 1 (LCN1) and transthyretin (TTR) showed lower and cystatin C higher abundance in COPD subjects; MSMB (PSP94) also shows slightly higher abundance in the COPD group, but in our study the dominant effect was an increase in both smoker groups (COPD and CS).

The comparison across all study groups indicated that the sputum levels generally exhibited a gradient from NS/FS to CS to COPD subjects; that is, proteins that were already overabundant in asymptomatic current smokers were further increased in COPD smokers (Fig. 5C). Furthermore, this observation of a general amplification of the CS effects in COPD subjects for the sputum proteome was also supported by a more global regression model in which the observed fold-changes for CS vs. NS and COPD vs. NS were compared and demonstrated a slope significantly larger than one (Fig. 5D).

The transcriptomics analysis of the cellular sputum fraction did not reveal any significantly differentially expressed mRNAs between the COPD and CS groups (Fig. 2B). Furthermore, a direct mRNA fold-change comparison for CS vs. NS and COPD vs. NS showed a trend toward global reduction of the CS effects in COPD subjects (Fig. 5E). Despite this global trend, several mRNA changes remained close to the diagonal, i.e. were only slightly different between CS and COPD. Similarly, a direct comparison of the mRNA abundances for the three identified main effects in CS, xenobiotic/oxidative stress response, M1 phenotype suppression, and M2 phenotype induction, did not show any clear trends between the CS and COPD groups (Fig. 5F). Interestingly, whereas the NPA analysis further supported that there is little change in

xenobiotic metabolism between CS and COPD, a slight weakening in macrophage activation comparing CS vs. NS and COPD vs. NS was observed (Fig. 5G). Specifically, nodes associated with M1 polarization (e.g., IFNG and the NFkB complex) showed less down-regulation in COPD vs. NS than CS vs. NS (Fig. 5H), and a similar trend was supported by gene set enrichment analysis (Supplementary Fig. 7). With this, the previously observed (weak) trend of a further immune-cell polarization in COPD (from M1 toward M2) [60], was not found in our study, which can likely be explained by the different subject and cell populations in our study.

In addition, we can cautiously interpret the deviations of the sputum results from the macrophages values obtained from the GEO:GSE13896 dataset by sputum-specific effects, such as mechanisms taking place in neutrophils, for which appropriate models are also available [47]. Fig. 5I shows that the backbone nodes contained in the *neutrophil chemotaxis* network model were on average higher in the COPD vs. NS than in the CS vs. NS pairwise comparison. Given the very simple structure of this network model that is causally consistent [45] and that contains almost exclusively activation edges, these observations indicate a significant increase of the chemotactic activity of the sputum neutrophils, confirmed by NPA network-level calculation (not shown) and exemplified by the positive value of the leukotriene B4 receptor (LTB4R) node [72]. With this, despite the weakness of the signal at individual gene-level, our network analysis was able to detect suggestive mechanisms accompanying the onset of COPD.

In summary, the sputum proteome reflected differences between current asymptomatic smokers and smokers with COPD. Strikingly, it was possible to distinguish COPD and CS subjects based on the proteomics data with a similar accuracy as based on the combination of three lung-function COPD metrics (FEV_1 pred., TLCO % predicted, and total COPD score) (Supplementary Fig. 9). In contrast, mRNA expression levels in sputum cells were similar between COPD and CS for the identified main biological effects and globally the mRNA differences were dampened rather than exaggerated in COPD vs. CS.

4. Discussion

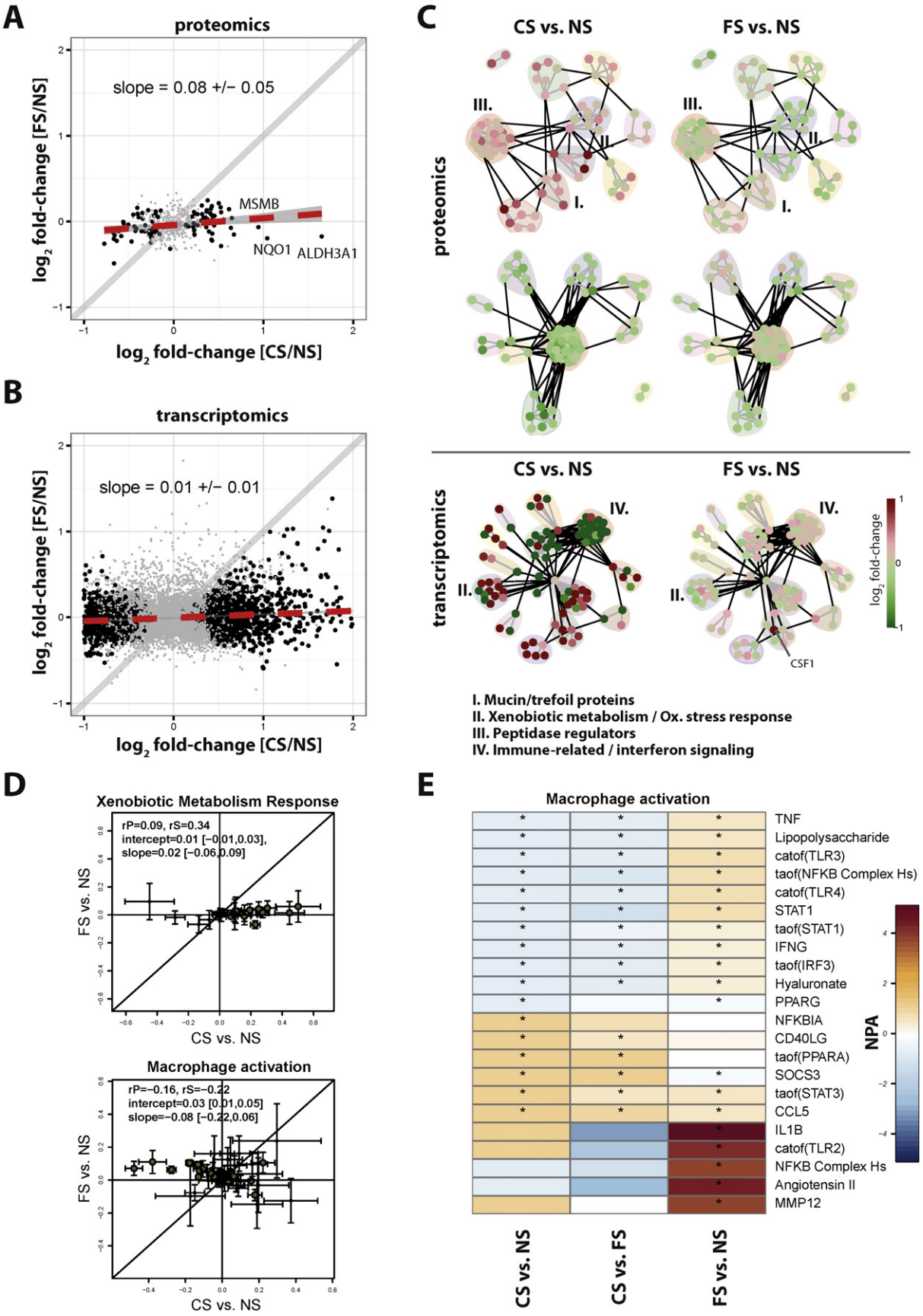
In this study, we employed a parallel-group case-controlled study design to assess how the free sputum proteome and the cellular sputum transcriptome reflected cigarette smoke exposure, smoking cessation, and are affected by the presence of early-stage COPD. While – with this study design– we did not directly follow the individual exposure response and disease onset, we evaluated to what degree the alterations observed in the sputum samples overall supported a general toxicant-exposure-to-disease-transition model for COPD (Fig. 6). This model is based on the observation that COPD initiation clearly is facilitated by long-term cigarette smoke exposure, smoking cessation is the most effective measure to reduce the risk of COPD development, and observable biological responses in asymptomatic smokers (e.g., increases in oxidative stress and inflammation in the lung) intersect with the identified COPD disease mechanisms that eventually result in chronic inflammation and structural lung damage [3]. The lungs of asymptomatic smokers still have sufficient biological buffering capacity to, for example, cope with the oxidative and xenobiotic challenges of cigarette-smoke exposure — effects such as up-regulation of oxidative stress response proteins are observable, but still reflect the physiological and reversible stress response. However, in many cases this homeostatic state cannot be maintained and accumulating structural damage initiates the manifestation of the disease. For this, the transition point will depend on genetic susceptibilities and on time-dependent processes (e.g., development of cellular senescence) that affect the potential of the tissue to handle the insult [18,19].

The sputum proteome and transcriptome clearly reflected some of the main biological effects of cigarette smoke exposure previously seen in other studies. Prominent xenobiotic metabolism and oxidative stress responses were detected for both sputum fractions as would be expected for cigarette smoke exposure [73,74]. For example, one of these proteins, ALDH3A1, recently was reported to be up-regulated in epithelial lining fluid of asymptomatic vs. never-smokers and was clearly associated with smoking status [75]. Here, sputum ALDH3A1 protein levels supported the accurate prediction of the study subjects' smoking status. Interestingly, in the present study, the up-regulated transcripts included genes that encode several metabolic enzymes (e.g., glucose-6-phosphate dehydrogenase (G6PD) and phosphogluconate dehydrogenase (PGD)) that likely contribute to the regeneration of NADPH as an important component of the oxidative stress response [76,77]. Other relevant detected changes in the sputum proteomes of current smokers included a change in mucus production (e.g., up-regulation of MUC5AC and trefoil factors 1/3) and up-regulation of TIMP metallopeptidase inhibitor 1 (TIMP1).

The detection of several xenobiotic and oxidative stress enzymes in the sputum supernatant raised a question about their source, and especially, whether these enzymes could play an active role in the mucus layer as a first line of defense against xenobiotic and oxidative challenges. Normal epithelial lining fluid has been found to contain high levels of reduced glutathione, which are further increased in smokers, and it has been suggested that this may provide extracellular protection against oxidative challenges [78]. In addition, an important role for transferrin in epithelial lining fluid for protection against lipid peroxidation has been identified [79]. Moreover, extracellular superoxide dismutase was found to be increased in sputum samples of current smokers and COPD subjects [80]. ALDH3A1 is an abundant detoxification enzyme in human cornea, its potential role in extracellular detoxification reactions has been demonstrated for human saliva [81,82], and it has been detected recently as a component of the epithelial lining fluid (see above). Together these findings suggest the active contribution of these detoxification enzymes as a first line of defense in the mucus layer/epithelial lining fluid of the airways.

An interesting observation from the transcriptomics data was the change in the apparent polarization of the overall sputum cell population from an M1 toward an M2 phenotype (Fig. 3D, Supplementary Fig. 7). Such a polarization shift was reported by Shaykhiev et al. for alveolar macrophages from asymptomatic smokers [60], and it has been suggested that such a shift could affect disease susceptibility [83] including that for lung cancer [84]. Furthermore, our network analysis (Supplementary Fig. 5) supported the notion that this suppression of the M1 phenotype is associated with the deactivation of several upstream signaling mechanisms (e.g., STAT1, TNF, and TLR3/4 signaling), which likely results in the down-regulation of NFkB and IFNg signaling as major transcriptional regulators of the M1 phenotype [59].

Fig. 4. Effect of long-term smoking cessation on the sputum proteome and transcriptome. (A) Fold-change comparison between CS vs. NS and FS vs. NS for the quantified sputum proteins. Proteins with significant differential abundance for CS vs. NS are indicated by large black dots. A linear regression line is fitted for these proteins and the slope and 95% confidence interval are given (dashed, red line; gray confidence range). (B) Fold-change comparison between CS vs. NS and FS vs. NS for the sputum transcriptomics data. Other details are similar to those in (A). (C) Comparison of the identified functional clusters/biological effects for CS vs. NS and FS vs. NS for the sputum proteomics (top panel) and the sputum transcriptomics (bottom panel) data. The clustered association networks from Fig. 3 are overlaid with the log2 fold-changes of the proteins/mRNAs for the considered comparisons. (D) Network perturbation amplitude (NPA) scatter plot comparisons for CS vs. NS and FS vs. NS for the backbone nodes of the *xenobiotic metabolism response* and the *macrophage activation* networks. The error bars indicate the 95% confidence range of the NPA values. The Pearson (rP) and Spearman (rS) correlation coefficients and the intercept and slope of the linear regression model (including the 95% confidence range) are indicated. (E) NPA values for the backbone nodes of the macrophage activation network. Nodes with a significant NPA value in the sputum cell CS vs. NS, CS vs. FS, or FS vs. NS comparison are shown (rows). The NPA for these nodes is compared for the three study group contrasts (columns, see color key) and rescaled by the value for CS vs. NS. "*" denotes significance; p-value <0.05 for both *uncertainty* and *specificity* statistics [45].


A
proteomics
slope = 0.08 +/− 0.05
MSMB
NQO1
ALDH3A1
log2 fold-change [FS/NS]
log2 fold-change [CS/NS]
B
transcriptomics
slope = 0.01 +/− 0.01
log2 fold-change [FS/NS]
log2 fold-change [CS/NS]
C
CS vs. NS
FS vs. NS
proteomics
transcriptomics
I.
II.
III.
IV.
CSF1
log2 fold-change
I. Mucin/trefoil proteins
II. Xenobiotic metabolism / Ox. stress response
III. Peptidase regulators
IV. Immune-related / interferon signaling
D
Xenobiotic Metabolism Response
rP=0.09, rS=0.34
intercept=0.01 [−0.01,0.03],
slope=0.02 [−0.06,0.09]
FS vs. NS
CS vs. NS
Macrophage activation
rP=−0.16, rS=−0.22
intercept=0.03 [0.01,0.05]
slope=−0.08 [−0.22,0.06]
FS vs. NS
CS vs. NS
E
Macrophage activation
TNF
Lipopolysaccharide
catof(TLR3)
taof(NFKB Complex Hs)
catof(TLR4)
STAT1
taof(STAT1)
IFNG
taof(IRF3)
Hyaluronate
PPARG
NFKBIA
CD40LG
taof(PPARA)
SOCS3
taof(STAT3)
CCL5
IL1B
catof(TLR2)
NFKB Complex Hs
Angiotensin II
MMP12
NPA
CS vs. NS
CS vs. FS
FS vs. NS

Smoking cessation has been identified as the only effective measure to reduce COPD morbidity and progression [13,85]. In our study, the subjects in the FS group had quit smoking for at least 1 year and the majority (78%) had quit for at least 5 years. In this cessation group, the observable smoking exposure effects in the sputum proteome and transcriptome had largely approached the levels of the subjects in the NS group (Figs. 2A/B, 4, 5E), including the oxidative and xenobiotic stress response and the changes in mucin/trefoil proteins. The polarization changes in the sputum cell population had also largely reverted (Fig. 5E), but interestingly some signs of a persistent (opposite) immune cell polarization toward M1 with increased INFG signaling were noticeable (Fig. 4, Supplementary Fig. 7). Cessation studies reported in the literature often look at shorter time frames. For example, Wozniak et al. [86] investigated the effect of 1–3 months of smoking cessation on the oxidative stress parameters in COPD subjects and found that even three months of smoking abstinence had partly restored the oxidant–antioxidant balance in plasma and erythrocytes. In another study, plasma markers of oxidative stress were still elevated after 6 months, but not significantly different from non-smokers after 12 months of smoking cessation [66]. In line with our results, the cessation effect on inflammation is more complex. For example, Willemse et al. [87] investigated the effect of 1-year smoking cessation on airway inflammation and found perpetuation of inflammation after smoking cessation in COPD subjects and reduction of some aspects of inflammation in asymptomatic smokers.

Thirteen proteins had significant differential abundance in the sputum of early-stage COPD subjects compared to the CS group (Fig. 5, Table 2). The majority of these proteins already showed signs of a concordant regulation in the CS group compared with the NS group (Fig. 5C). We also found evidence that the effects observed for the sputum proteome of COPD subjects generally could be regarded as a (slight) amplification of the smoke exposure effects before disease developed (Fig. 5D). Interestingly, the sputum transcriptome did not reveal any significantly differentially regulated transcripts between the COPD and CS group and the identified biological effects were not increased in the COPD compared to the CS group; on the contrary, an opposite global trend in the mRNA expression levels was detected (Fig. 5E/F). These observations lend support to the model that early-stage COPD disease represents a manifestation of the chronic toxicant exposure effects (e.g., facilitated by shifts in stress response balances) rather than a distinct, separate disease entity (see above and Fig. 6).

Previous studies of the sputum proteome of COPD subjects include the 2D-gel based proteomics studies by Ohlmeier et al. and Nicholas et al. [23,25,88]. Ohlmeier et al. [25] identified polymeric immunoglobulin receptor (PIGR) as an up-regulated protein in asymptomatic smokers with further elevation in smokers with COPD disease. Whereas differential PIGR expression did not reach significance in our study after multiple-hypothesis correction, a tendency for increased abundance in COPD subjects was observed (p-value (COPD vs. CS) = 0.08, p-value (COPD vs. NS) = 0.04, Supplementary Figure S8). With a study by Nicholas et al. [23], significant down-regulation of the candidate biomarkers apolipoprotein A1 (APOA1), α-2-HS glycoprotein (ASHG), and albumin (ALB) were shared. In addition, down-regulation of APOA1, ALB, and transferrin also has been reported for lung biopsies of smokers [89]. Moreover, even for the proteins that did not reach significance in our study, the trends were similar to those reported by Nicholas et al. [23] (Supplementary Fig. 8); for example, for lipocalin-1 (LCN1), a down-regulation trend was observed in COPD subjects in our study (p-value (COPD vs. CS) = 0.08, p-value (COPD vs. NS) = 0.07). Of note, beta-microseminoprotein (MSMB, also known as PSP94) was identified as an up-regulated protein in all three previous studies [23,25,88], which may reflect mainly the current smoking status rather than the disease status. Previously, MSMB was suggested as a candidate biomarker for increased glandular activity and secretory/goblet cell hyperplasia in any type of airways disease [23,90].

In total, at least six of the 13 differentially abundant proteins in COPD vs. CS have been linked previously to COPD (Table 2). Of these proteins, the tissue inhibitor of matrix-metalloproteinases (TIMP1) showed the strongest correlation with lung function parameters (Fig. 5B). For example, Ziora et al. and Aaron et al. have reported elevated TIMP1 levels in COPD subjects compared with asymptomatic smokers [69,70]. BPIFB1 (LPLUNC1) is another example protein that was elevated in the COPD group. BPIFB1 family members have been found in airway submucosal glands and in a population of airway goblet cells [93], and a strong association between increased BPIFB1 expression and idiopathic pulmonary fibrosis has been reported [94,95]. Strikingly, overlapping localization was observed for BPIFB1 and MUC5B likely extending to "bronchiolized epithelium" of a COPD patient [94], which further supports its potential role in COPD. C6orf58 is an example for a protein without prior link to COPD, but little information about C6orf58 is available [91,92].

From a computational methodological point-of-view, it is worth noting several features of the NPA approach that we used to analyze the transcriptomics data [45,46]. First, it was applied to cell type-specific inflammatory network models such as *macrophage activation* and *neutrophil chemotaxis*[47], even though the cellular sputum RNA contained the mixed contributions from various cell types (mostly macrophages and neutrophils). We justified the step by comparing our results to several public datasets containing isolated alveolar macrophages, which displayed significant similarities and suggested in particular that the CS vs. NS and FS vs. NS pairwise comparisons could be understood in terms of macrophage-related processes, such as their different polarization states. The NPA approach also displayed high sensitivity in the case of the COPD vs. CS and FS vs. NS pairwise comparisons, where it enabled the detection of perturbed mechanisms even in the absence of differentially expressed genes (i.e., with adjusted p-values <0.05). For example, for the activity of the IFNG node, which was found to reverse its sign upon smoking cessation, the NPA results were confirmed by the more common GSEA approach (Supplementary Fig. 6), which provided further confidence in our systems biology-based approach.

Taken together, we and others found that induced sputum can provide a relevant window into the responsive changes of the airways to

Fig. 5. Sputum proteome shows a difference between COPD subjects and current smokers. (A) Fold-change heatmap for the 13 differentially expressed proteins between the COPD and CS groups. The log2 fold-change is color coded. "*" indicates statistically significant differential abundance (Benjamini–Hochberg adjusted p-value <0.05). (B) Correlation plots with FEV1% of predicted (top panel) and TLCO predicted (bottom panel) for the 13 differentially abundant proteins in COPD vs. CS. The bars show the Pearson correlation coefficient and the red lines the coefficient of determination (R2) for each protein. (C) Median relative protein abundance values across all groups compared with the never smoker group (NS) for the differentially abundant proteins in the COPD vs. CS comparison. Up-regulated proteins are in red, down-regulated proteins are in blue, the median of the up- and down-regulated group is indicated by a black dashed line. (D) Fold-change comparison between CS vs. NS and COPD vs. NS for the quantified sputum proteins. Proteins with significant differential abundance for CS vs. NS are indicated by a large black dot. A linear regression line is fitted for these proteins and the slope and 95% confidence interval are given (dashed, red line; gray confidence range). (E) Fold-change comparison between CS vs. NS and COPD vs. NS for the quantified sputum transcriptomics data. Other details are similar to (D). (F) Median relative mRNA expression values across all groups compared with the never smoker group (NS) for the differentially abundant proteins in the COPD vs. CS comparison for three gene groups, M1-related genes, M2-related genes (see Fig. 3D), and genes in the xenobiotic/oxidative stress cluster (see Fig. 3B). Other details are similar to (C). (G) Network perturbation amplitude (NPA) scatter plot based on transcriptomics data for CS vs. NS and COPD vs. NS for the backbone nodes of the *xenobiotic metabolism response* and the *macrophage activation* networks. The error bars indicate the 95% confidence range of the NPA values. The Pearson (rP) and Spearman (rS) correlation coefficients and the intercept and slope of the linear regression model (including the 95% confidence range) are indicated. (H) NPA for the macrophage activation network for the sputum transcriptomics data and the study on alveolar macrophages by Shaykhiev et al. (GSE13896) [60]. Nodes from the macrophage activation network with a significant NPA value in the sputum cell CS vs. NS comparison are shown (rows). The NPA for these nodes is compared across the study groups for both studies (columns, see color key) and rescaled by the respective CS vs. NS value. "*" denotes significance; p-value <0.05. (I) NPA scatter plot comparisons for CS vs. NS and COPD vs. NS for the backbone nodes of the *neutrophil chemotaxis* network (see panel G for details).

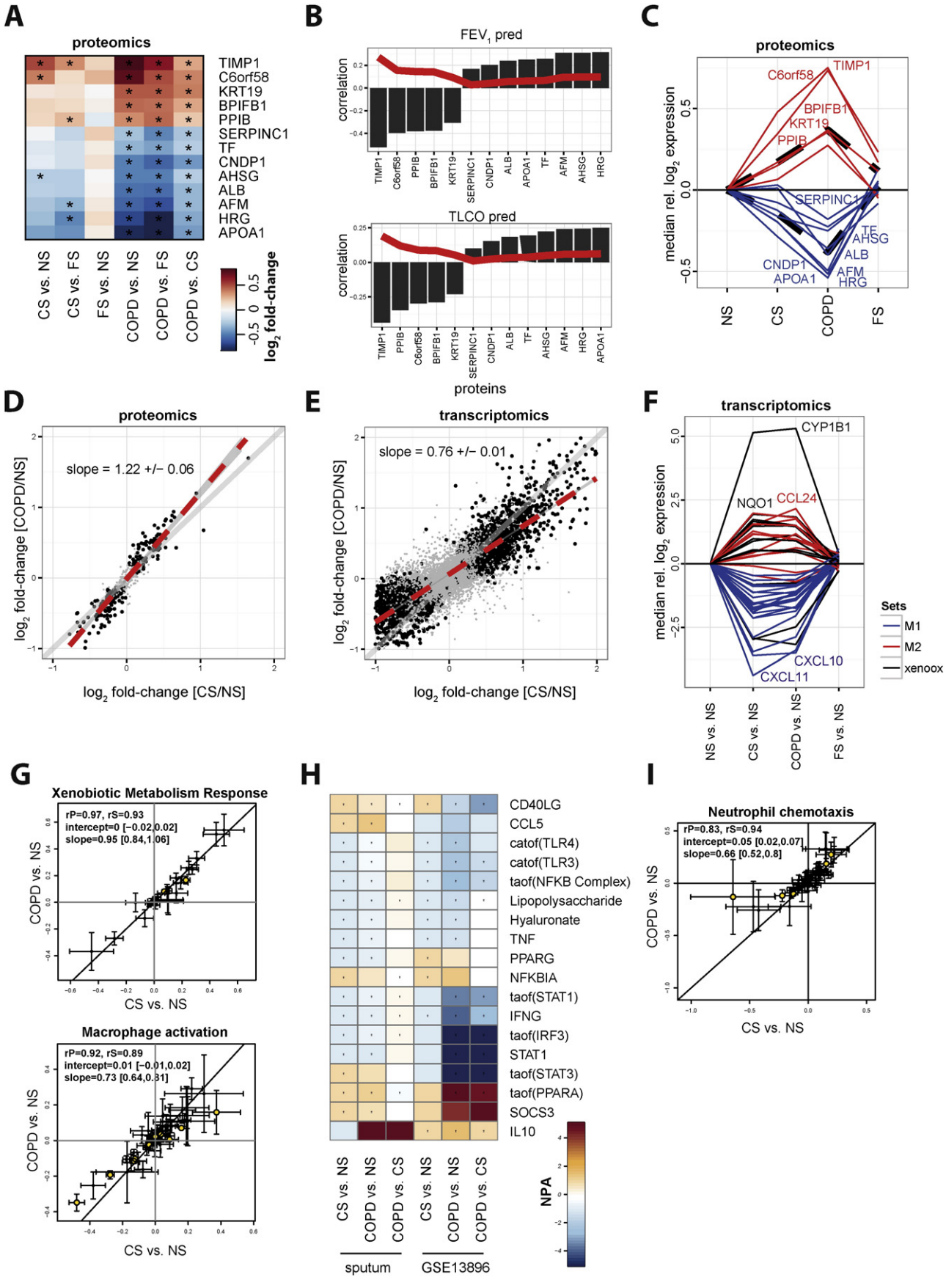

A
proteomics
TIMP1
C6orf58
KRT19
BPIFB1
PPIB
SERPINC1
TF
CNDP1
AHSG
ALB
AFM
HRG
APOA1
CS vs. NS
CS vs. FS
FS vs. NS
COPD vs. NS
COPD vs. FS
COPD vs. CS
log2 fold-change
B
FEV1 pred
correlation
TLCO pred
proteins
C
proteomics
median rel. log2 expression
NS
CS
COPD
FS
D
proteomics
slope = 1.22 +/− 0.06
log2 fold-change [COPD/NS]
log2 fold-change [CS/NS]
E
transcriptomics
slope = 0.76 +/− 0.01
F
transcriptomics
CYP1B1
NQO1
CCL24
CXCL10
CXCL11
Sets
M1
M2
xenoox
NS vs. NS
CS vs. NS
COPD vs. NS
FS vs. NS
G
Xenobiotic Metabolism Response
rP=0.97, rS=0.93
intercept=0 [−0.02,0.02]
slope=0.95 [0.84,1.06]
COPD vs. NS
CS vs. NS
Macrophage activation
rP=0.92, rS=0.89
intercept=0.01 [−0.01,0.02]
slope=0.73 [0.64,0.81]
H
CD40LG
CCL5
catof(TLR4)
catof(TLR3)
taof(NFKB Complex)
Lipopolysaccharide
Hyaluronate
TNF
PPARG
NFKBIA
taof(STAT1)
IFNG
taof(IRF3)
STAT1
taof(STAT3)
taof(PPARA)
SOCS3
IL10
NPA
sputum
GSE13896
I
Neutrophil chemotaxis
rP=0.83, rS=0.94
intercept=0.05 [0.02,0.07]
slope=0.66 [0.52,0.8]

Table 2
Differentially abundant proteins between early-stage COPD subjects and asymptomatic current smokers.

UniProtKB accession	Gene symbol	Description	Log2 FC	FDR	Literature
P08727	**KRT19**	Keratin, type I cytoskeletal 19	0.38	0.025	↑ BALF, PF [98] ↑ BALF, PF [99]
Q6P5S2	**C6orf58**	UPF0762 protein C6orf58	0.37	0.031	
Q5H9A7	**TIMP1**	Metalloproteinase inhibitor 1	0.29	0.016	↑ **IS, COPD**[69] ↑ **IS, COPD**[70]
Q8TDL5	**BPIFB1 [LPLUNC1]**	BPI fold-containing family B member 1	0.28	0.028	↑ LB, CF [100] ↑ LP, COPD [101]
P23284	**PPIB**	Peptidyl-prolyl cis-trans isomerase B (cyclophilin B)	0.21	0.016	↑ NL, Asthma [102]
P02787	**TF**	Serotransferrin	−0.23	0.016	↓ BALF, COPD [103] ↑ **IS, COPD**[25]
P02765	**AHSG**	Alpha-2-HS-glycoprotein	−0.24	0.025	↓ **IS, COPD**[23]
P01008	**SERPINC1**	Antithrombin-III	−0.27	0.021	
P43652	**AFM**	Afamin	−0.32	0.018	
P02768	**ALB**	Serum albumin	−0.32	0.007	↓ **IS, COPD**[23] ↓ **IS, COPD**[71]
P04196	**HRG**	Histidine-rich glycoprotein	−0.36	0.016	
P02647	**APOA1**	Apolipoprotein A-I	−0.39	0.016	↓ IS, COPD [23]
Q96KN2	**CNDP1**	Beta-Ala-His dipeptidase	−0.42	0.016	

FC, fold-change [COPD vs. CS] in our study; FDR, false discovery rate (FDR)-adjusted p-value; IS, induced sputum; BALF, bronchio-alveolar lavage; NL, nasal lavage; LP, lung parenchyma; PF, pulmonary fibrosis; ↑↓ = up-/down-regulated; bold, findings for sputum of COPD subjects.

toxicant exposures and potentially lung diseases [21,96]. However, it should be noted that sputum represents a challenging sample type, in part because of the abundance of highly glycosylated proteins and variations in sample dilution. Because of these dilution effects, it is now recognized that data normalization is especially critical for sputum samples. For example, shifts in cell type distributions rather than absolute cell numbers have been found to be more relevant [55]. Similarly, normalization based on the total protein content was used in our and previous sputum proteomics studies, and a recent validation study for COPD biomarkers provided examples that are useful for this type of normalization [71].

While induced sputum clearly captures the responsive changes of the airways to toxicant exposure, the question is to which extent disease specific alterations can be assessed. In our study, the detected changes in the sputum of COPD subjects generally represented an amplification of the trends already observed in the CS controls, which is in line with the findings of previous proteomic studies [23,25]. Similarly, Gao et al. [97] discussed the challenges associated with distinguishing COPD subjects from asymptomatic smokers compared with distinguishing asymptomatic or diseased smokers from non-smoker controls. Therefore, it will be pertinent to explore other sampling sources, especially to capture the earliest signs of disease onset. For example, it has been suggested recently that the earliest lesions in COPD are driven by airway epithelial basal progenitor cells [64], which could encourage the development of more specific assays for targeting COPD changes in populations of this cell type.

In conclusion, the present study was designed to assess how biological changes are reflected in sputum samples over the whole course of the "toxicant-exposure-to-disease-transition model" for COPD (Fig. 6). We found that the sputum proteome and transcriptome both captured the mostly reversible biological responses of the airways to cigarette smoke exposure including the oxidative/xenobiotic stress response, changes in mucus production, and protease balance, and alterations in the state of the respiratory immune system (Fig. 3). The transition from a still physiological toxicant exposure response to early-stage COPD disease leaves its marks in the sputum proteome, but the changes are not strong and mostly an augmentation of the already established exposure effects. While a few promising targets for COPD detection

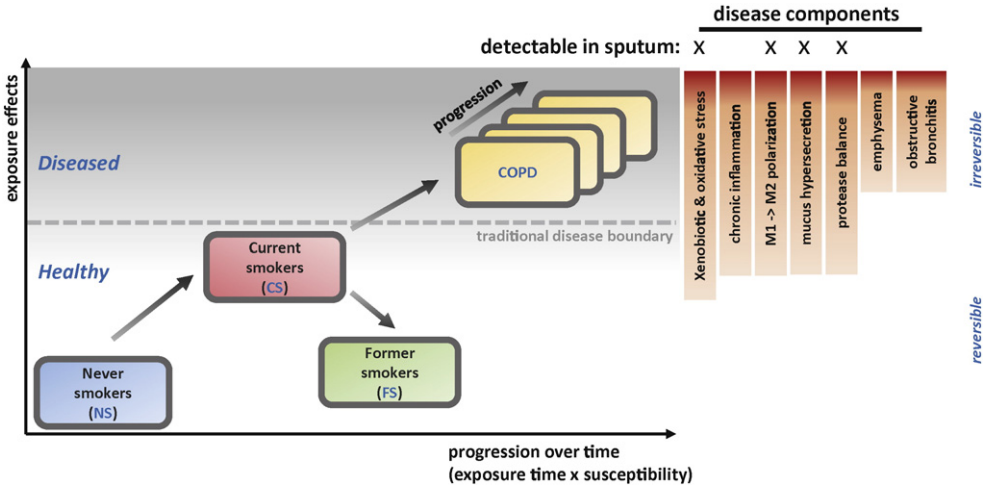


Fig. 6. Schematic representation of a basic toxicant-exposure-to-disease-transition model for COPD. This model represents the four main states (never smokers, current smoker, former smoker, COPD) and their dominant transitions for smoking-caused COPD. Cigarette smoke exposure induces several effects (Y-axis), of which a subset is detectable in sputum samples. These effects often appear amplified and become irreversible in COPD with a not necessarily strict healthy-disease boundary. Overall, disease development is facilitated by cigarette exposure over time and affected by disease susceptibility of the individual (X-axis) (see text). Smoking cessation is the best established measure to reduce exposure effects and to prevent the transition from a current smoker to COPD. See Sections 1 and 4 for more details.

have been identified, the identification of more specific signs of early disease likely will require the discovery of novel mechanistic targets, potentially those linked to early alterations in basal progenitor cells.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2015.08.009.

Transparency Document

The Transparency document associated with this article can be found, in the online version.

Acknowledgments

We thank the Heart Lung Centre at the Queen Anne Street Medical Centre for executing the clinical trial. Sam Ansari conducted the data submission of the transcriptomics data.

References

[1] E. Diaz-Guzman, D.M. Mannino, Epidemiology and prevalence of chronic obstructive pulmonary disease, Clin. Chest Med. 35 (2014) 7–16.
[2] Organization WH, The 10 Leading Causes of Death in the World, 2000 and 2012, 2014.
[3] J. Vestbo, S.S. Hurd, A.G. Agusti, P.W. Jones, C. Vogelmeier, A. Anzueto, et al., Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary, Am. J. Respir. Crit. Care Med. 187 (2013) 347–365.
[4] P. Shirtcliffe, M. Weatherall, J. Travers, R. Beasley, The multiple dimensions of airways disease: targeting treatment to clinical phenotypes, Curr. Opin. Pulm. Med. 17 (2011) 72–78.
[5] D.R. Gold, X. Wang, D. Wypij, F.E. Speizer, J.H. Ware, D.W. Dockery, Effects of cigarette smoking on lung function in adolescent boys and girls, N. Engl. J. Med. 335 (1996) 931–937.
[6] A. Langhammer, R. Johnsen, A. Gulsvik, T.L. Holmen, L. Bjermer, Sex differences in lung vulnerability to tobacco smoking, Eur. Respir. J. 21 (2003) 1017–1023.
[7] R. Hooper, P. Burney, W.M. Vollmer, M.A. McBurnie, T. Gislason, W.C. Tan, et al., Risk factors for COPD spirometrically defined from the lower limit of normal in the BOLD project, Eur. Respir. J. 39 (2012) 1343–1353.
[8] J. Vestbo, A. Agusti, E.F. Wouters, P. Bakke, P.M. Calverley, B. Celli, et al., Should we view chronic obstructive pulmonary disease differently after ECLIPSE? A clinical perspective from the study team, Am. J. Respir. Crit. Care Med. 189 (2014) 1022–1030.
[9] A. Løkke, P. Lange, H. Scharling, P. Fabricius, J. Vestbo, Developing COPD: a 25 year follow up study of the general population, Thorax 61 (2006) 935–939.
[10] R. Perez-Padilla, A. Schilmann, H. Riojas-Rodriguez, Respiratory health effects of indoor air pollution, Int. J. Tuberc. Lung Dis. 14 (2010) 1079–1086.
[11] J. Zhang, K.R. Smith, Indoor air pollution: a global health concern, Br. Med. Bull. 68 (2003) 209–225.
[12] G. Bettoncelli, F. Blasi, V. Brusasco, S. Centanni, A. Corrado, F. De Benedetto, et al., The clinical and integrated management of COPD. An official document of AIMAR (Interdisciplinary Association for Research in Lung Disease), AIPO (Italian Association of Hospital Pulmonologists), SIMER (Italian Society of Respiratory Medicine), SIMG (Italian Society of General Medicine), Multidiscip. Respir. Med. 9 (2014) 25.
[13] N. Godtfredsen, T. Lam, T. Hansel, M. Leon, N. Gray, C. Dresler, et al., COPD-related morbidity and mortality after smoking cessation: status of the evidence, Eur. Respir. J. 32 (2008) 844–853.
[14] R.M. Tuder, I. Petrache, Pathogenesis of chronic obstructive pulmonary disease, J. Clin. Invest. 122 (2012) 2749–2755.
[15] J. Domagala-Kulawik, Effects of cigarette smoke on the lung and systemic immunity, J. Physiol. Pharmacol. 59 (2008) 19–34.
[16] I. Rahman, S.K. Biswas, A. Kode, Oxidant and antioxidant balance in the airways and airway diseases, Eur. J. Pharmacol. 533 (2006) 222–239.
[17] P. Maestrelli, L. Richeldi, M. Moretti, L. Fabbri, Analysis of sputum in COPD, Thorax 56 (2001) 420–422.
[18] A. Bar-Shai, A. Sagiv, R. Alon, V. Krizhanovsky, The role of Clara cell senescence in the pathogenesis of COPD, Eur. Respir. J. 44 (2014) 3245.
[19] M. Kumar, W. Seeger, R. Voswinckel, Senescence-associated secretory phenotype and its possible role in chronic obstructive pulmonary disease, Am. J. Respir. Cell Mol. Biol. 51 (2014) 323–333.
[20] S. Haenen, E. Clynen, B. Nemery, P.H. Hoet, J.A. Vanoirbeek, Biomarker discovery in asthma and COPD: application of proteomics techniques in human and mice, EuPA Open Proteomics 4 (2014) 101–112.
[21] G. Pelaia, R. Terracciano, A. Vatrella, L. Gallelli, M.T. Busceti, C. Calabrese, et al., Application of proteomics and peptidomics to COPD, BioMed. Res. Int. 2014 (2014) 764581.
[22] B. Nicholas, P. Skipp, R. Mould, S. Rennard, D. Davies, C. O'Connor, et al., Shotgun proteomic analysis of human-induced sputum, Proteomics 6 (2006) 4390.
[23] B.L. Nicholas, P. Skipp, S. Barton, D. Singh, D. Bagmane, R. Mould, et al., Identification of lipocalin and apolipoprotein A1 as biomarkers of chronic obstructive pulmonary disease, Am. J. Respir. Crit. Care Med. 181 (2010) 1049–1060.
[24] R.D. Gray, G. MacGregor, D. Noble, M. Imrie, M. Dewar, A.C. Boyd, et al., Sputum proteomics in inflammatory and suppurative respiratory diseases, Am. J. Respir. Crit. Care Med. 178 (2008) 444.
[25] S. Ohlmeier, W. Mazur, A. Linja-Aho, N. Louhelainen, M. Ronty, T. Toljamo, et al., Sputum proteomics identifies elevated PIGR levels in smokers and mild-to-moderate COPD, J. Proteome Res. 11 (2012) 599–608.
[26] D. Singh, S.M. Fox, R. Tal-Singer, J. Plumb, S. Bates, P. Broad, et al., Induced sputum genes associated with spirometric and radiological disease severity in COPD ex-smokers, Thorax 66 (2011) 489–495.
[27] J. Vestbo, W. Anderson, H.O. Coxson, C. Crim, F. Dawber, L. Edwards, et al., Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE), Eur. Respir. J. 31 (2008) 869–873.
[28] C.D. Kelstrup, C. Young, R. Lavallee, M.L. Nielsen, J.V. Olsen, Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer, J. Proteome Res. 11 (2012) 3487–3497.
[29] R Development Core Team, R: A Language and Environment for Statistical Computing, 2009.
[30] N.A. Karp, W. Huber, P.G. Sadowski, P.D. Charles, S.V. Hester, K.S. Lilley, Addressing accuracy and precision issues in iTRAQ quantitation, Mol. Cell. Proteomics 9 (2010) 1885–1897.
[31] W. Huber, A. Von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression, Bioinformatics 18 (2002) S96–S104.
[32] S.M. Herbrich, R.N. Cole, K.P. West, K. Schulze, J.D. Yager, J.D. Groopman, et al., Statistical Inference from multiple iTRAQ experiments without using common reference standards, J. Proteome Res. 12 (2012) 594–604.
[33] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al., Bioconductor: open software development for computational biology and bioinformatics, Genome Biol. 5 (2004) R80.
[34] J.A. Vizcaino, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Rios, et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination, Nat. Biotechnol. 32 (2014) 223–226.
[35] Drăghici, Sorin, Statistics and data analysis for microarrays using R and bioconductor, CRC Press, 2011 See: https://www.crcpress.com/Statistics-and-Data-Analysis-for-Microarrays-Using-R-and-Bioconductor/Drghici/9781439809754.
[36] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al., Bioconductor: open software development for computational biology and bioinformatics, Genome Biol. 5 (2004) R80.
[37] R.A. Irizarry, B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, et al., Exploration, normalization, and summaries of high density oligonucleotide array probe level data, Biostatistics 4 (2003) 249–264.
[38] A. Kauffmann, R. Gentleman, W. Huber, arrayQualityMetrics—a bioconductor package for quality assessment of microarray data, Bioinformatics 25 (2009) 415–416.
[39] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, et al., STRING v9.1: protein–protein interaction networks, with increased coverage and integration, Nucleic Acids Res. 41 (2013) D808–D815.
[40] M. Rosvall, C. Bergstrom, Maps of information flow reveal community structure in complex networks, Proceedings of the National Academy of Sciences USA, Citeseer, 2007.
[41] J. Chen, E.E. Bardes, B.J. Aronow, A.G. Jegga, ToppGene suite for gene list enrichment analysis and candidate gene prioritization, Nucleic Acids Res. 37 (2009) W305–W311.
[42] H. Fang, J. Gough, The 'dnet' approach promotes emerging research on cancer patient survival, Genome Med. 6 (2014) 64.
[43] G. Csardi, T. Nepusz, The igraph software package for complex network research, Int. J. Complex Syst. 1695 (2006) 1–9.
[44] R Development Core Team, R: A Language and Environment for Statistical Computing, 2007.
[45] F. Martin, T.M. Thomson, A. Sewer, D.A. Drubin, C. Mathis, D. Weisensee, et al., Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks, BMC Syst. Biol. 6 (2012) 54.
[46] F. Martin, A. Sewer, M. Talikka, Y. Xiang, J. Hoeng, M.C. Peitsch, Quantification of biological network perturbations for mechanistic insight and diagnostics using two-layer causal models, BMC Bioinforma. 15 (2014) 238.
[47] J.W. Westra, W.K. Schlage, A. Hengstermann, S. Gebel, C. Mathis, T. Thomson, et al., A modular cell-type focused inflammatory process network model for non-diseased pulmonary tissue, Bioinforma. Biol. Insights 7 (2013) 167.
[48] W.K. Schlage, J.W. Westra, S. Gebel, N.L. Catlett, C. Mathis, B.P. Frushour, et al., A computable cellular stress network model for non-diseased pulmonary and cardiovascular tissue, BMC Syst. Biol. 5 (2011) 168.
[49] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.
[50] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Methodol. (1996) 267–288.
[51] A.L. Tarca, M. Lauria, M. Unger, E. Bilal, S. Boue, K.K. Dey, et al., Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge, Bioinformatics 29 (2013) 2892–2899.
[52] Characterisation of COPD heterogeneity in the ECLIPSE cohort. A. Agusti, P.M. Calverley, B. Celli, H.O. Coxson, L.D. Edwards, D.A. Lomas, W. MacNee, B.E. Miller, S. Rennard, E.K. Silverman, R. Tal-Singer, E. Wouters, J.C. Yates, J. Vestbo, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) investigators, Respir Res. 11 (2010 Sep 10) 122, http://dx.doi.org/10.1186/1465-9921-11-122.

[53] R. O'Donnell, C. Peebles, J. Ward, A. Daraker, G. Angco, P. Broberg, et al., Relationship between peripheral airway dysfunction, airway obstruction, and neutrophilic inflammation in COPD, Thorax 59 (2004) 837.
[54] V.M. Keatings, P.D. Collins, D.M. Scott, P.J. Barnes, Differences in interleukin-8 and tumor necrosis factor-alpha in induced sputum from patients with chronic obstructive pulmonary disease or asthma, Am. J. Respir. Crit. Care Med. 153 (1996) 530–534.
[55] D. Singh, L. Edwards, R. Tal-Singer, S. Rennard, Research Sputum neutrophils as a Biomarker in COPD: Findings from the ECLIPSE Study, 2010.
[56] B.M. Fischer, E. Pavlisko, J.A. Voynow, Pathogenic triad in COPD: oxidative stress, protease–antiprotease imbalance, and inflammation, Int. J. Chron. Obstruct. Pulm. Dis. 6 (2011) 413–421.
[57] B.M. Fischer, J.A. Voynow, A.J. Ghio, COPD: balancing oxidants and antioxidants, Int. J. Chron. Obstruct. Pulm. Dis. 10 (2015) 261–276.
[58] J.V. Fahy, B.F. Dickey, Airway mucus function and dysfunction, N. Engl. J. Med. 363 (2010) 2233–2247.
[59] A. Sica, A. Mantovani, Macrophage plasticity and polarization: in vivo veritas, J. Clin. Invest. 122 (2012) 787–795.
[60] R. Shaykhiev, A. Krause, J. Salit, Y. Strulovici-Barel, B.-G. Harvey, T.P. O'Connor, et al., Smoking-dependent reprogramming of alveolar macrophage polarization: implication for pathogenesis of chronic obstructive pulmonary disease, J. Immunol. 183 (2009) 2867–2883.
[61] J.W. Graff, L.S. Powers, A.M. Dickson, J. Kim, A.C. Reisetter, I.H. Hassan, et al., Cigarette smoking decreases global microRNA expression in human alveolar macrophages, PLoS One 7 (2012) e44066.
[62] H. Vanni, A. Kazeros, R. Wang, B.G. Harvey, B. Ferris, B.P. De, et al., Cigarette smoking induces overexpression of a fat-depleting gene AZGP1 in the human, Chest 135 (2009) 1197–1208.
[63] A. Cabral, P. Voskamp, A.M. Cleton-Jansen, A. South, D. Nizetic, C. Backendorf, Structural organization and regulation of the small proline-rich family of cornified envelope precursors suggest a role in adaptive barrier function, J. Biol. Chem. 276 (2001) 19231–19237.
[64] R. Shaykhiev, R.G. Crystal, Early events in the pathogenesis of chronic obstructive pulmonary disease. Smoking-induced reprogramming of airway epithelial basal progenitor cells, Ann. Am. Thorac. Soc. 11 (Suppl. 5) (2014) S252–S258.
[65] O. Auerbach, J.B. Gere, J.B. Forman, T.G. Petrick, H.J. Smolin, G.E. Muehsam, et al., Changes in the bronchial epithelium in relation to smoking and cancer of the lung, CA Cancer J. Clin. 8 (1958) 53–56.
[66] J.F. Zhou, X.F. Yan, F.Z. Guo, N.Y. Sun, Z.J. Qian, D.Y. Ding, Effects of cigarette smoking and smoking cessation on plasma constituents and enzyme activities related to oxidative stress, Biomed. Environ. Sci. 13 (2000) 44–55.
[67] B.W.M. Willemse, D.S. Postma, W. Timens, N.H.T. Ten Hacken, The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation, Eur. Respir. J. 23 (2004) 464–476.
[68] J. Ebbert, P. Yang, C. Vachon, R. Vierkant, J. Cerhan, A. Folsom, et al., Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women, J. Clin. Oncol. 21 (2003) 921–926.
[69] D. Ziora, S. Dworniczak, J. Kozielski, Induced sputum metalloproteinases and their inhibitors in relation to exhaled nitrogen oxide and sputum nitric oxides and other inflammatory cytokines in patients with chronic obstructive pulmonary disease, J. Physiol. Pharmacol. 59 (Suppl. 6) (2008) 809–817.
[70] S.D. Aaron, K.L. Vandemheen, T. Ramsay, C. Zhang, Z. Avnur, T. Nikolcheva, et al., Multi analyte profiling and variability of inflammatory markers in blood and induced sputum in patients with stable COPD, Respir. Res. 11 (2010) 41.
[71] S. Ropcke, O. Holz, G. Lauer, M. Muller, S. Rittinghausen, P. Ernst, et al., Repeatability of and relationship between potential COPD biomarkers in bronchoalveolar lavage, bronchial biopsies, serum, and induced sputum, PLoS One 7 (2012) e46207.
[72] J.L. Corhay, M. Henket, D. Nguyen, B. Duysinx, J. Sele, R. Louis, Leukotriene B4 contributes to exhaled breath condensate and sputum neutrophil chemotaxis in COPD, Chest 136 (2009) 1047–1054.
[73] L. Zuo, F. He, G.G. Sergakis, M.S. Koozehchian, J.N. Stimpfl, Y. Rong, et al., Interrelated role of cigarette smoking, oxidative stress, and immune response in COPD and corresponding treatments, Am. J. Physiol. Lung Cell. Mol. Physiol. 307 (2014) L205–L218.
[74] P.A. Kirkham, P.J. Barnes, Oxidative stress in COPD, Chest 144 (2013) 266–273.
[75] L. Franciosi, D.S. Postma, M. van den Berge, N. Govorukhina, P.L. Horvatovich, F. Fusetti, et al., Susceptibility to COPD: differential proteomic profiling after acute smoking, PLoS One 9 (2014) e102037.
[76] A.R. Agarwal, L. Zhao, H. Sancheti, I.K. Sundar, I. Rahman, E. Cadenas, Short-term cigarette smoke exposure induces reversible changes in energy metabolism and cellular redox status independent of inflammatory responses in mouse lungs, Am. J. Physiol. Lung Cell. Mol. Physiol. 303 (2012) L889–L898.
[77] M. Wamelink, E. Struys, C. Jakobs, The biochemistry, metabolism and inherited defects of the pentose phosphate pathway: a review, J. Inherit. Metab. Dis. 31 (2008) 703–717.
[78] A. Cantin, S. North, R. Hubbard, R. Crystal, Normal alveolar epithelial lining fluid contains high levels of glutathione, J. Appl. Physiol. 63 (1987) 152–157.
[79] E.R. Pacht, W.B. Davis, Role of transferrin and ceruloplasmin in antioxidant activity of lung epithelial lining fluid, J. Appl. Physiol. 64 (1988) 2092–2099.
[80] E.A. Regan, W. Mazur, E. Meoni, T. Toljamo, J. Millar, K. Vuopala, et al., Smoking and COPD increase sputum levels of extracellular superoxide dismutase, Free Radic. Biol. Med. 51 (2011) 726–732.
[81] M. Bogucka, J. Giebultowicz, K. Zawada, P. Wroczynski, J. Wierzchowski, M. Pietrzak, et al., The oxidation status of ALDH3A1 in human saliva and its correlation with antioxidant capacity measured by ORAC method, Acta Pol. Pharm. 66 (2009) 477–482.
[82] J. Giebultowicz, M. Dziadek, P. Wroczynski, K. Woznicka, B. Wojno, M. Pietrzak, et al., Salivary aldehyde dehydrogenase — temporal and population variability, correlations with drinking and smoking habits and activity towards aldehydes contained in food, Acta Biochim. Pol. 57 (2010) 361–368.
[83] M.R. Stämpfli, G.P. Anderson, How cigarette smoke skews immune responses to promote infection, lung disease and cancer, Nat. Rev. Immunol. 9 (2009) 377–384.
[84] J. Ma, L. Liu, G. Che, N. Yu, F. Dai, Z. You, The M1 form of tumor-associated macrophages in non-small cell lung cancer is positively associated with survival time, BMC Cancer 10 (2010) 112.
[85] K.F. Rabe, S. Hurd, A. Anzueto, P.J. Barnes, S.A. Buist, P. Calverley, et al., Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary, Am. J. Respir. Crit. Care Med. 176 (2007) 532–555.
[86] A. Wozniak, D. Gorecki, M. Szpinda, C. Mila-Kierzenkowska, B. Wozniak, Oxidant–antioxidant balance in the blood of patients with chronic obstructive pulmonary disease after smoking cessation, Oxidative Med. Cell. Longev. 2013 (2013) 897075.
[87] B.W. Willemse, N.H. ten Hacken, B. Rutgers, I.G. Lesman-Leegte, D.S. Postma, W. Timens, Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers, Eur. Respir. J. 26 (2005) 835–845.
[88] S. Ohlmeier, M. Vuolanto, T. Toljamo, K. Vuopala, K. Salmenkivi, M. Myllarniemi, et al., Proteomics of human lung tissue identifies surfactant protein A as a marker of chronic obstructive pulmonary disease, J. Proteome Res. 7 (2008) 5125–5132.
[89] S.G. Kelsen, X. Duan, R. Ji, O. Perez, C. Liu, S. Merali, Cigarette smoke induces an unfolded protein response in the human lung: a proteomic approach, Am. J. Respir. Cell Mol. Biol. 38 (2008) 541–550.
[90] A. Tilley, M. Staudt, J. Fuller, B. De, R. Crystal, N. Qadir, Smoking-induced upregulation of microseminoprotein beta gene expression in the human airway, Am. J. Respir. Crit. Care Med. 185 (2012) A6052.
[91] C. Chang, M. Hu, Z. Zhu, L.J. Lo, J. Chen, J. Peng, Liver-enriched gene 1a and 1b encode novel secretory proteins essential for normal liver development in zebrafish, PLoS One 6 (2011) e22910.
[92] A.K. Enjapoori, T.R. Grant, S.C. Nicol, C.M. Lefèvre, K.R. Nicholas, J.A. Sharp, Monotreme lactation protein is highly expressed in monotreme milk and provides antimicrobial protection, Genome Biol. Evol. 6 (2014) 2754–2773.
[93] C.D. Bingle, K. Wilson, H. Lunn, F.A. Barnes, A.S. High, W.A. Wallace, et al., Human LPLUNC1 is a secreted product of goblet cells and minor glands of the respiratory and upper aerodigestive tracts, Histochem. Cell Biol. 133 (2010) 505–515.
[94] C.D. Bingle, B. Araujo, W.A. Wallace, N. Hirani, L. Bingle, What is top of the charts? BPIFB1/LPLUNC1 localises to the bronchiolised epithelium in the honeycomb cysts in UIP, Thorax 68 (2013) 1167–1168.
[95] I.V. Yang, C.D. Coldren, S.M. Leach, M.A. Seibold, E. Murphy, J. Lin, et al., Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis, Thorax 68 (12) (2013) 1114–1121, http://dx.doi.org/10.1136/thoraxjnl-2012-202943.
[96] J.G. Shaw, A. Vaughan, A.G. Dent, P.E. O'Hare, F. Goh, R.V. Bowman, et al., Biomarkers of progression of chronic obstructive pulmonary disease (COPD), J. Thorac. Dis. 6 (2014) 1532–1547.
[97] J. Gao, H. Ilumets, A. Linja-aho, S. Ohlmeier, N. Ishikawa, H. Kobayashi, et al., Sputum biomarkers of inflammation in smokers and subjects with COPD, Eur. Respir. J. 42 (2013) P841.
[98] M. Inage, H. Nakamura, S. Kato, H. Saito, S. Abe, T. Hino, et al., Levels of cytokeratin 19 fragments in bronchoalveolar lavage fluid correlate to the intensity of neutrophil and eosinophil-alveolitis in patients with idiopathic pulmonary fibrosis, Respir. Med. 94 (2000) 155–160.
[99] N. Dobashi, J. Fujita, Y. Ohtsuki, I. Yamadori, T. Yoshinouchi, T. Kamei, et al., Elevated serum and BAL cytokeratin 19 fragment in pulmonary fibrosis and acute interstitial pneumonia, Eur. Respir. J. 14 (1999) 574–578.
[100] L. Bingle, K. Wilson, M. Musa, B. Araujo, D. Rassl, W.A. Wallace, et al., BPIFB1 (LPLUNC1) is upregulated in cystic fibrosis lung disease, Histochem. Cell Biol. 138 (2012) 749–758.
[101] Y. Liu, J. Latoche, J. Pilewski, Y. Di, Differential expression of SPLUNC1 And LPLUNC1 in lungs of healthy human subjects and COPD patients, Am. J. Respir. Crit. Care Med.: Am. Thorac. Soc. 181 (2010) A2915–A2915.
[102] E.J. Stemmy, A.S. Benton, J. Lerner, S. Alcala, S.L. Constant, R.J. Freishtat, Extracellular cyclophilin levels associate with parameters of asthma in phenotypic clusters, J. Asthma 48 (2011) 986–993.
[103] S.W. Stites, M.E. Nelson, L.J. Wesselius, Transferrin concentrations in serum and lower respiratory tract fluid of mechanically ventilated patients with COPD or ARDS, Chest 107 (1995) 1681–1685.